概率论

第一册　概率论基础

复旦大学　编

高等学校试用教材

概　率　论

第一册　概率论基础

复旦大学　编

高等教育出版社

内 容 提 要

全书分为三册：概率论基础、数理统计、随机过程。本书材料比较丰富，并配备一定数量的习题。其内容经适当选择后，可作为综合大学和师范院校数学专业试用教材，也可供理科其他有关专业及工科某些专业作为教学参考书。

高等学校试用教材

概 率 论

第一册 概率论基础

复 旦 大 学 编

*

高等教育出版社出版

新华书店北京发行所发行

天水新华印刷厂印装

*

开本787×1092 1/32 印张 10.25 字数 247,000

1979年4月第1版 1987年3月第9次印刷

印数 171,951—183,780

书号 13010·0225 定价 1.45元

目　录

序言……1
第一章　事件与概率……6
§1. 随机现象与统计规律性……6
§2. 样本空间与事件……13
§3. 古典概型……21
§4. 几何概率……36
§5. 概率空间……42
第一章小结……53
习题……54
第二章　条件概率与统计独立性……59
§1. 条件概率，全概率公式，贝叶斯公式……59
§2. 事件独立性……67
§3. 贝努里试验与直线上的随机游动……75
§4. 二项分布与普阿松分布……87
第二章小结……100
习题……102
第三章　随机变量与分布函数……107
§1. 随机变量及其分布……107
§2. 随机向量，随机变量的独立性……129
§3. 随机变量的函数及其分布……142
第三章小结……157
习题……159
第四章　数字特征与特征函数……164
§1. 数学期望，方差，矩……164
* §2. 熵与信息……184
* §3. 母函数……197
§4. 特征函数……

* §5. 多元正态分布……215
第四章小结……226
习题……227
第五章　极限定理……233
§1. 贝努里试验场合的极限定理……233
§2. 收敛性……252
§3. 独立同分布场合的极限定理……270
* §4. 强大数定律……278
* §5. 中心极限定理……294
第五章小结……306
习题……307
附录一　常用分布表……316
附录二　普阿松分布 $P\{\xi=r\}=\frac{\lambda^r}{r!}e^{-\lambda}$ 的数值表　正态分布密度函数 $\varphi(x)=\frac{1}{\sqrt{2\pi}}e^{-x^2/2}$ 及分布函数 $\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-t^2/2}dt$ 的数值表……320

序　言

概率论是从数量侧面研究随机现象规律性的数学学科，理论严谨，应用广泛，发展迅速.

概率论是数学的一个有特色的分支．一方面，它有别开生面的研究课题，有自己独特的概念和方法，内容丰富，结果深刻；另一方面，它与其它数学分支又有紧密的联系，它是近代数学的重要组成部分.

目前，概率论的理论与方法已广泛应用于工业、农业、军事和科学技术中．如预测和滤波应用于空间技术和自动控制；时间序列分析应用于石油勘探和经济管理；马尔可夫过程与点过程应用于地震预报和气象预报；数理统计方法应用于工农业生产等等.在理论联系实际方面，概率论是数学最活跃的分支之一.

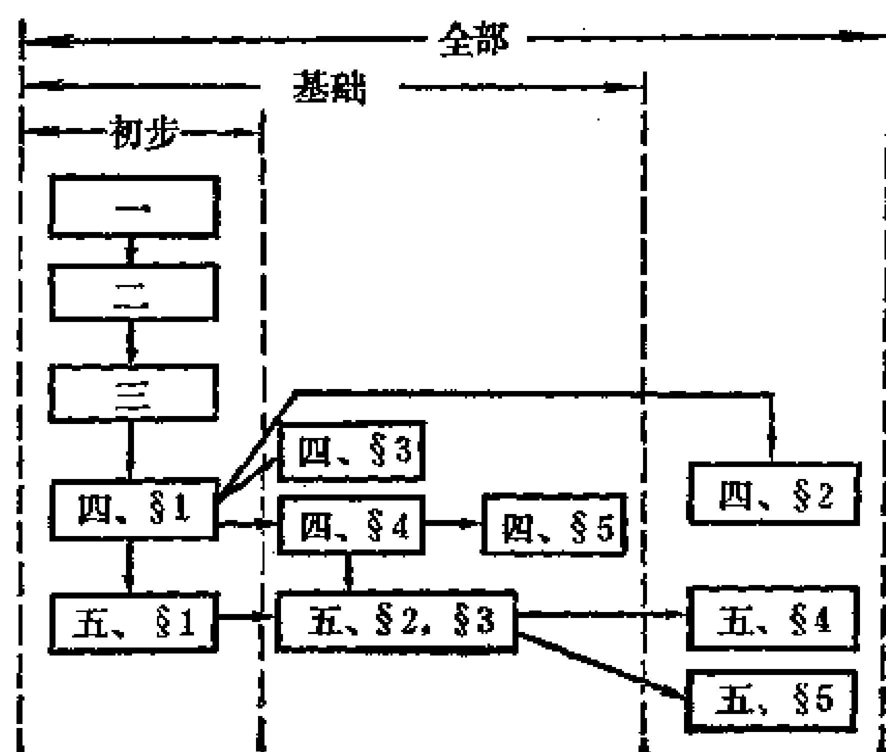
概率论的理论和方法向各个基础学科、工程学科的渗透，是近代科学技术发展的特征之一．概率论与其它学科相结合发展成不少边缘学科，如生物统计，统计物理和数学地质等；它又是许多新的重要学科的基础，如信息论，控制论，可靠性理论和人工智能等.

今天，高等学校理科、工科和医科的许多专业的教学计划中都设置了概率论或数理统计课程，以使学生初步掌握处理随机现象的基本理论和方法，培养他们解决某些有关实际问题的能力．编写本书的目的就在于为上述课程提供教材或基本教学参考书.

全书共分三部分：概率论基础、数理统计和随机过程，分三册出版．作为一个整体，全书的内容是连贯的，体例是一致的，它包含了概率论学科的主要基础知识；同时这三部分的内容又有自己的独立性，因此通过其他课本学过概率论基础的读者，可以直接学习本书的数理统计部分或随机过程部分．全书的内容基本上假定读者具有微积分的基础知识，数理统计部分还要求读者懂得线性代数，随机过程部分有些章节还需要一些实变函数论和泛函分析的初步知识．

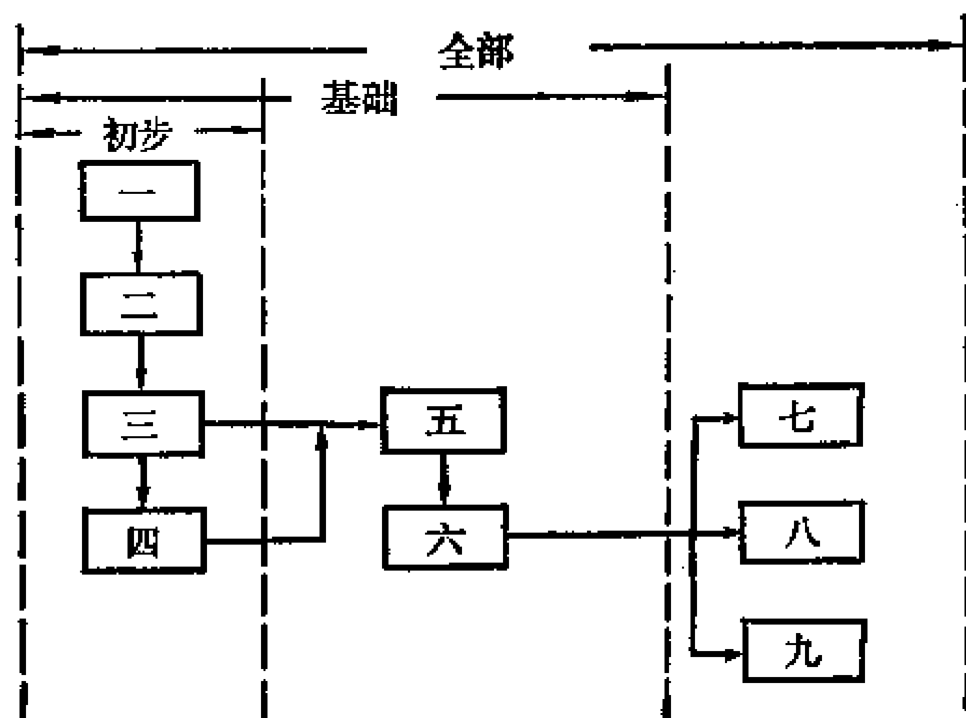
全书内容的组织是"模块式"的，作者希望通过适当的选择，本书将能适应多方面的教学需要．三部分内部各章节的逻辑关系如下①：

I　概率论基础

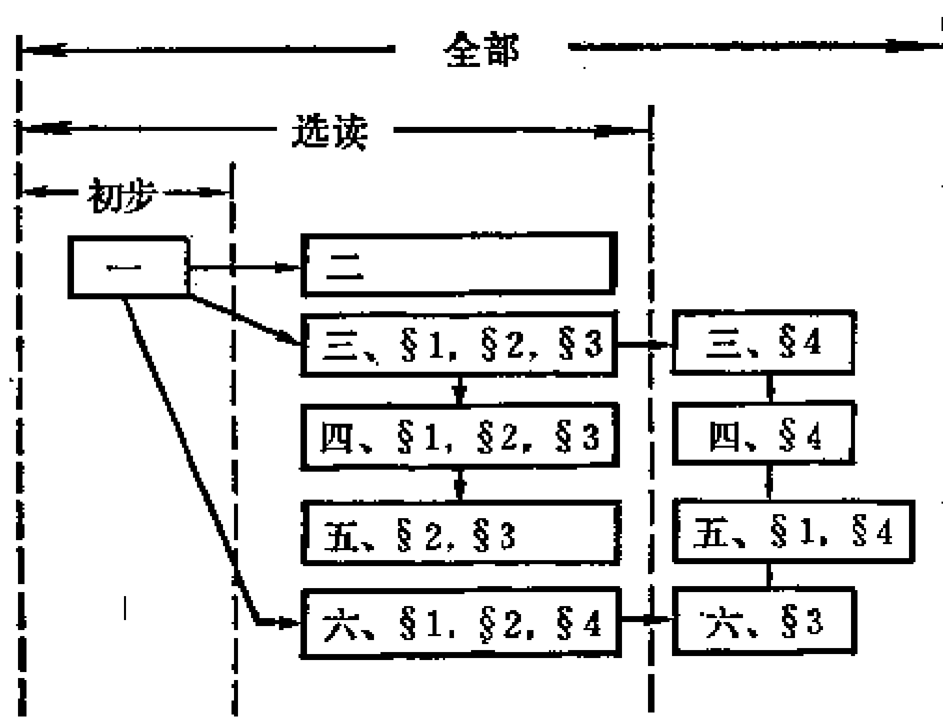


① 表内中文数字(如一，二等)为章标号．

Ⅱ 数理统计



Ⅲ 随机过程



从教学时数来看，下面也提供几种可供选择的安排①：

① 表内数字（如36，48 等）为学时数。

	概率论基础	数理统计	随机过程
初步	36	24～36	10
基础	48	50	
选读			(一，二) 30
			(一，三，四，五) 50
			(一，六) 20
全读	60	约 100	80

本书中的定理、公式和图表等均按节编号，引用同节的公式时，只指明编号；同章跨节引用时还指出节号，例如(2.21)表示同章第2节(21)式；跨章引用时还指明章节号，例如(3.2.21)表示第三章第2节(21)式.

标有星号的章、节、段有相对独立性，跳过它们不损害全书的连贯.

本书的编写工作在吴立德同志的领导下进行，结构与体例由编写者集体讨论商定. 概率论基础部分由李贤平同志编写；数理统计部分一至八章由卞国瑞同志编写，第九章由汪嘉冈同志编写；随机过程部分第一、二、三、六章由吴立德同志编写，第四、五章由汪嘉冈同志编写. 最后，汪嘉冈同志审阅了全书.

在编写过程中，我们得到了郑绍濂同志的指导和帮助. 本书初稿于1978年秋季完成后，上海师范大学魏宗舒教授等曾对本书提出了许多宝贵的意见和建议；本校徐业基同志和吴立鹏同志阅读过部分原稿并提出不少意见和建议，谨此一并致谢.

1979年初在广州举行了有14个兄弟院校26名代表参加的审稿会议，与会代表认真阅读了全书，并对全书的结构、选材和细节等方面提出了十分中肯的意见和建议，在此基础上我们又

书作了修改．利用这个机会，我们向主审单位北京大学以及参加审稿会议的其它兄弟院校同志表示衷心的感谢．

由于水平所限，编写工作又较匆促，因此一定存在不少缺点和错误，欢迎读者批评、指正．

编者

1979 年 1 月

第一章　事件与概率

§1. 随机现象与统计规律性

一、随机现象

概率论是研究随机现象的数量规律的数学分支．为了说明什么是随机现象，让我们先来看一个例子．设某车间有二百台车床，由于经常需要检修、测量、调换刀具、变换位置等种种原因，因此，即使在生产期间，各台车床还是时常需要停车．若每台车床有百分之六十的时间在开动，而每台车床开动时需要耗电1千瓦，问应供给这个车间多少电力才能保证此车间正常生产？

类似的例子在许多实际问题中出现，解决这类问题当然具有重大意义．

显然，若供给这个车间200千瓦的电能则此车间便能正常生产．但这样做不合算，因为每台车床的开工率只有60%，也就是说，平均起来这个车间中同时在工作的车床只有120台，供给它们200千瓦的电能太多．供给120千瓦的电能行吗？这又太少些，因为有时工作的车床数会超过120台，若只供给120千瓦，则这时会出现因电力不足而使车床无法正常运转，那么到底供给多少电才能既保证生产正常进行而又节约电力呢？

用概率论的方法能给这个问题以完满的解决，现在我们仅把结果列出，详细的解法以后再讲．计算表明，只要供给这个车间141千瓦的电就够了，虽然在这时也可能碰到因电力不足而不能正常生产的情况．但这种机会很少，它小于0.1%，即在8小时工

作中一般只有半分钟会碰到这种情况，这显然影响不大。但节约出来的59千瓦电能却能用来做许多别的用途。

解决这个问题的关键在于要计算出某时刻同时工作着的车床数，但是由于某台车床在某时刻是否开工很难预先确定，这受到许多偶然因素的影响，即工作着的车床数是一个受许多偶然因素影响的量，对于这种量，在概率论以外的数学分支中还没有得到研究。

原来，在自然界和人类社会中存在着两类不同的现象。

当我们多次观察自然现象和社会现象后，会发现许多事情在一定的条件下必然会发生。例如在没有外力作用的条件下，作等速直线运动的物体必然继续作等速直线运动。又如在标准大气压下，水加热到100℃时必然会沸腾等等。这种在一定条件下，必然会发生的事情称为**必然事件**。反之，那种在一定条件下，必然不会发生的事情就称为**不可能事件**。例如在不受外力作用的条件下，作等速直线运动的物体改变其等速直线运动状态是不可能的。

从所举例子中看出，必然事件和不可能事件，虽然形式相反，但是两者的实质是相同的。必然事件的反面就是不可能事件，而不可能事件的反面就是必然事件。

所有这种现象我们称之为**决定性现象**，它广泛地存在于自然现象和社会现象中，概率论以外的数学分支研究的是决定性现象的数量规律。

但是在自然现象和社会现象中也还广泛存在着与决定性现象有着本质区别的另一类现象，上述车间供电问题就是一例。

类似的例子还可以举出很多，例如用同一仪器多次测量同一物体的重量，所得结果彼此总是略有差异，这是由于诸如测量仪器受大气影响，观察者生理上或心理上的变化等等偶然因素引起的。同样地，同一门炮向同一目标发射多发同种炮弹，弹落点也不

一样，因为炮弹制造时种种偶然因素对炮弹质量有影响，此外，炮筒位置的误差，天气条件的微小变化等等都影响弹落点．再如从某生产线上用同一种工艺生产出来的灯泡的寿命也有差异等等．总之所举这些现象的一个共同的特点是：在基本条件不变的情况下，一系列试验或观察会得到不同的结果．换句话说，就个别的试验或观察而言，它会时而出现这种结果，时而出现那种结果，呈现出一种偶然性．这种现象称为**随机现象**．对于随机现象通常关心的是在试验或观察中某个结果是否出现，这些结果称为随机事件，简称**事件**．例如过马路交叉口时可能遇上各种颜色的交通指挥灯，这是一个随机现象，而"遇到红灯"则是一个随机事件．以后我们用 $A, B, C, \cdots\cdots$ 等大写拉丁字母表示随机事件．

二、频率稳定性

正如恩格斯所指出的："在表面上是偶然性在起作用的地方，这种偶然性始终是受内部的隐蔽着的规律支配的，而问题只是在于发现这些规律．"（恩格斯：《路德维希·费尔巴哈和德国古典哲学的终结》，人民出版社，1972 年，第 38 页）．

人们经过长期的实践发现，虽然个别随机事件在某次试验或观察中可以出现也可以不出现，但在大量试验中它却呈现出明显的规律性——频率稳定性．

对于随机事件 A，若在 N 次试验中出现了 n 次，则称

$$F_N(A)=\frac{n}{N}$$

为随机事件在 N 次试验中出现的**频率**．

在掷一枚硬币时，既可能出现正面，也可能出现反面，预先作出确定的判断是不可能的，但是假如硬币均匀，直观上出现正面与出现反面的机会应该相等，即在大量试验中出现正面的频率，应接近于 50%，为了验证这点，历史上曾有不少人做过这个试验，其结

果如下①:

实验者	掷硬币次数	出现正面次数	频率
蒲丰	4040	2048	0.5069
皮尔逊	12000	6019	0.5016
皮尔逊	24000	12012	0.5005

又如,在英语中某些字母出现的频率远远高于另外一些字母。在进行了更深入的研究之后,人们还发现各个字母被使用的频率相当稳定。例如,下面就是英文字母使用频率的一份统计表②。其他各种文字也都有着类似的规律.

字母	空格	E	T	O	A	N	I	R	S
频率	0.2	0.105	0.072	0.0654	0.063	0.059	0.055	0.054	0.052
字母	H	D	L	C	F	U	M	P	Y
频率	0.047	0.035	0.029	0.023	0.0225	0.0225	0.021	0.0175	0.012
字母	W	G	B	V	K	X	J	Q	Z
频率	0.012	0.011	0.0105	0.008	0.003	0.002	0.001	0.001	0.001

字母使用频率的研究,对于打字机键盘的设计(在方便的地方安排使用频率较高的字母键)、印刷铅字的铸造(使用频率高的应铸得多些)、信息的编码(常用字母用较短的码)、密码的破译等等方面都是十分有用的.

另一个验证频率稳定性的著名试验是由英国生物统计学家高尔顿(Galton)设计的。它的试验模型如图1所示.

自上端放入一小球,任其自由下落,在下落过程中当小球碰到

① 引自格涅坚科《概率论教程》,高等教育出版社,第44页.

② 引自L. Brillouin, Science and Information Theory, New York, 1956.

钉子时，从左边落下与从右边落下的机会相等．碰到下一排钉子时又是如此．最后落入底板中的某一格子．因此，任意放入一球，则此球落入那一个格子，预先难以确定．但是实验证明，如放入大量小球，则其最后所呈现的曲线，几乎总是一样的．也就是说，小球落入各个格子的频率十分稳定．这个试验模型称为**高尔顿板**．这个试验中呈现出来的规律性，在学习第五章极限定理之后，就会有更深刻的理解．

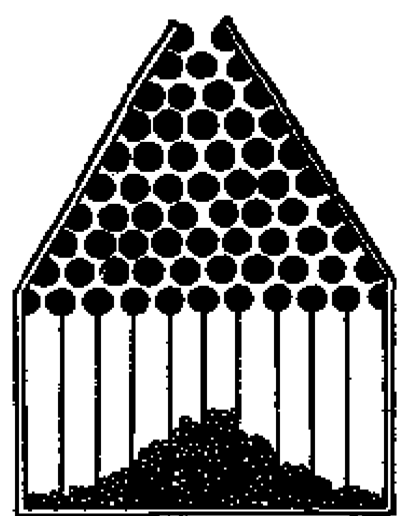


图 1　高尔顿板

同样，如果多次测量同一物体，其结果虽略有差异，但当测量次数增加时，就会越来越清楚地呈现出一些规律性：测量值的平均值在某固定常数附近波动，诸测量值在此常数两旁的分布呈现某种对称性．又如在射击的例子中，当射击次数不多时，炮弹的弹落点似乎是前后左右杂乱无章，看不出什么明显的规律；但当射击次数增加时，弹落点的分布就会呈现出一定的规律性：如弹落点关于目标的分布略呈对称性，偏离目标远的弹落点比偏离目标近的弹落点少等等．其他如灯泡寿命等，在进行多次观察或试验后，也都可以发现类似的规律性．

上述种种事实表明，随机现象有其偶然性的一面，也有其必然性的一面．这种必然性表现为大量试验中随机事件出现的频率的稳定性，即一个随机事件出现的频率常在某个固定的常数附近摆动，这种规律性我们称之为**统计规律性**．频率的稳定性说明随机事件发生的可能性大小是随机事件本身固有的、不随人们意志而改变的一种客观属性，因此可以对它进行度量．

对于一个随机事件 A，用一个数 $P(A)$ 来表示该事件发生的可能性大小，这个数 $P(A)$ 就称为随机事件 A 的**概率**．因此概率度量

了随机事件发生的可能性的大小.

对于随机现象，光讨论它可能出现什么结果，价值不大，而指出各种结果出现的可能性的大小则具有很大意义. 有了概率的概念就使我们能对随机现象进行定量研究，由此建立了一个新的数学分支——概率论.

三、频率与概率

既然概率 $P(A)$ 度量了随机事件 A 发生的可能性大小，可以预料，在 N 次重复试验中，若 $P(A)$ 较大，则频率 $F_N(A)=\frac{n}{N}$ 也较大. 反之若 $P(A)$ 很小，则 $F_N(A)$ 也很小，而且概率 $P(A)$ 应与频率有许多相似的性质. 以下我们先对频率的性质进行一番考察.

首先频率具有非负性

$$F_N(A)\geqslant 0 \tag{1}$$

其次，对于必然发生的事件，在 N 次试验中应出现 N 次. 若以 Ω 记必然事件，则应有

$$F_N(\Omega)=1 \tag{2}$$

还有，若 A 及 B 是两个不会同时发生的随机事件，以 $A+B$ 表示 A 或 B 至少出现其一这个事件，则应有

$$F_N(A+B)=F_N(A)+F_N(B) \tag{3}$$

这个性质称为**频率的可加性**.

当然还可以列出频率的许多性质，但上述三个性质是最基本的. 例如，“不可能事件在 N 次试验中出现的频率为 0”，“任何随机事件在 N 次试验中出现的频率不大于 1”，“对于有限个两两不会同时发生的随机事件也有频率可加性”，这些性质都可以由(1)，(2)，(3)推出.

最后，根据上述频率稳定性的讨论似乎可以提出这样的猜想，即当 N 足够大时 $F_N(A)$ 与 $P(A)$ 应充分接近. 这一点有很大的启

发性，在历史上它一直是概率论研究的一个重大课题．以后我们将会看到，在很一般的条件下，这个结论的确成立，但同时还须对问题的提法进一步明确化．

频率与概率的上述关系有时还提供了求某事件概率的一种手段，即当N足够大时，用它的频率来作为概率的近似值．以后我们将会看到，这种做法大有好处．

四、概率论简史

概率论是一门研究随机现象数量规律的学科．早在十七世纪中叶便开始了对随机现象的研究，在这些研究中建立了概率论的一些基本概念如事件，概率，随机变量，数学期望等．当时研究的模型较简单，就是现在通称的古典概型．

其后，随着生产实践的发展，特别是在射击理论、人寿保险、测量误差等工作中提出的一些概率问题，促使人们在概率论的极限定理方面进行深入研究．起初主要对贝努里(Bernoulli)试验概型进行，其后则推广到更为一般的场合．极限定理的研究在十八世纪和十九世纪整整200年中成了概率论研究的中心课题．在本世纪初，由于新的更有力的数学方法的引入，这些问题得到了较好解决．

虽然概率论历史悠久，但是它的严格的数学基础的建立以及理论研究和实际应用的极大发展却主要是本世纪的事情．

由于物理学(如统计物理)、生物学以及工程技术（如自动电话、无线电技术)发展的推动，概率论得到了飞速的发展．理论课题不断扩大与深入，概率论的思想渗入各个学科成了近代科学发展明显的特征之一．

由于各个数学分支的发展与互相渗透，概率论的严格的数学基础被建立起来，古典问题得到了解决，新的概念和工具不断出现，概率论成了数学的一个活跃分支．

概率论大大地扩大了它的应用范围. 特别在最近几十年中，概率论的方法被引入各个工程技术学科和社会学科. 目前，概率论在近代物理，无线电与自动控制，工厂产品的质量控制，农业试验，公用事业等等方面都找到了重要应用，这些实际需要也有力地推动了概率论的新发展，有些还形成了边缘学科（如信息论、排队论）.

在这个时期内，由于生物学和农业试验的推动，数理统计学也获得了很大发展，它以概率论为理论基础又为概率论应用提供了有力的工具，两者互相推动，迅速发展. 而概率论本身的研究则转入以随机过程为中心课题，取得了许多理论上和应用上都有重要价值的结果.

§2. 样本空间与事件

一、样本空间

从本节开始，我们将逐步引进概率论的基本概念. 样本空间与事件是最基本的两个概念.

对随机现象的研究必然要联系到对客观的事物进行"观察"或"试验". 我们假定，这种"观察"与"试验"可以在相同条件下重复进行.

我们感兴趣的是试验的结果. 例如掷一次硬币，我们关心的是出现正面或出现反面，这是两个可能出现的结果. 假如我们考察的是掷二次硬币的试验，则可能出现的结果有(正，正)，(正，反)，(反，正)，(反，反)四种；如果掷三次硬币，则结果还要复杂，但还是可以把它们描述出来. 总之，为了研究随机试验，首先需要知道这个试验可能出现的结果，这些结果称为**样本点**，一般用ω表示，样本点全体构成**样本空间**，用Ω表示. 在具体问题中，给定样本空间是描述随机现象的第一步.

下面举一些例子.

[例 1] 在研究英文字母使用情况时，把样本空间选为 $\Omega=\{$空格, A, B, …, Z$\}$是适宜的，这个样本空间只有有限个样本点，是比较简单的样本空间.

[例 2] 观察一小时中落在地球上某一区域的宇宙射线数，可能的结果一定是非负整数，而且很难指定一个数作为它的上界，这样，可以把样本空间取为 $\Omega=\{0, 1, 2, \cdots\}$. 这个样本空间含有无穷多个样本点，但这些样本点可以依照某种次序排列出来，以后我们将称它的点数为可列个.

[例 3] 讨论某地区的气温时，我们自然把样本空间取为 $\Omega=(-\infty, \infty)$，或 $\Omega=[a, b]$. 这个样本空间包含有无穷多个样本点，它们充满一个区间，不是一个可列集.

[例 4] 考察地震震源时，可以把样本点取为(x, y, z)，其中 x 表示震源的经度，y 表示纬度，z 表示深度. 这时，样本空间是三维空间中某一区域.

可见，随着问题不同，样本空间可以相当简单，也可以相当复杂.

在今后讨论中，经常把样本空间认为是预先给定的. 当然对于一个实际问题或一个随机现象，如何用一个恰当的样本空间来描述它也很值得研究. 但是在概率论的研究中，一般都认为样本空间是给定的，这是必要的抽象，这种抽象使我们能更好地把握住随机现象的本质，而且得到的结果能广泛地应用. 事实上，一个样本空间可以概括各种实际内容大不相同的问题；例如只包含两个样本点的样本空间既能作为掷硬币出现正、反面的模型，也能用于产品检验中出现“好品”及“废品”，又能用于气象中“下雨”与“不下雨”，以及公用事业排队现象中“有人排队”与“无人排队”等等. 尽管问题的实际内容如此不同，但有时却能归结为相同的概率模型.

我们后面常以摸球等作为例子也是由于这个原因，它能使问题的本质更为突出.

二、事件

有了样本空间的概念，就可以定义事件. 我们还是从考察一个例子开始.

［例 5］ 某袋中装有 4 只白球和 2 只黑球，我们考虑依次从中摸出两球所可能出现的事件. 若对球进行编号，4 只白球分别编为 1,2,3,4 号，2 只黑球编为 5,6 号. 如果用数对(i, j)表示第一次摸得 i 号球，第二次摸得 j 号球，则可能出现的结果是

$$\begin{array}{l}(1,2),(1,3),(1,4),(1,5),(1,6)\\(2,1),(2,3),(2,4),(2,5),(2,6)\\(3,1),(3,2),(3,4),(3,5),(3,6)\\(4,1),(4,2),(4,3),(4,5),(4,6)\\(5,1),(5,2),(5,3),(5,4),(5,6)\\(6,1),(6,2),(6,3),(6,4),(6,5)\end{array} \qquad (*)$$

把这 30 个结果作为样本点，则构成了样本空间. 在这个问题中，这些样本点是我们感兴趣的事件；但是我们也可以研究下面另外一些事件：

A：第一次摸出黑球；

B：第二次摸出黑球；

C：第一次及第二次都摸出黑球.

后面这些事件与前面那些事件的不同处在于这些事件是可以分解的，例如为了 A 出现必须而且只须下列样本点之一出现：

$$\begin{array}{l}(5,1),(5,2),(5,3),(5,4),(5,6)\\(6,1),(6,2),(6,3),(6,4),(6,5)\end{array}$$

前面的 30 个事件是由单个样本点构成；后面这三个事件，每一个事件都是由若干个样本点构成，总之，它们都是样本点的某个

集合.

所谓给定一个点的集合 S,是指对于任何一个点ω, 都可以确定它是不是属于 S. 如果是,则记为 $\omega\in S$; 如果不是,则记为 $\omega\notin S$. 按照这种定义,单个点也是一个点集,习惯上还约定不包含任何点的集合也是一个点集,称为空集,一般记为$\varnothing$.

今后,我们把事件定义为样本点的某个集合,称某事件发生当且仅当它所包含的某一个样本点出现.

我们把样本空间Ω也作为一个事件, 因为在每次试验中必然出现Ω中的某个样本点, 也即Ω必然发生, 所以常称Ω为**必然事件**. 类似地,我们把空集$\varnothing$也作为一个事件,它在每次试验中都不会发生,称为**不可能事件**.

必然事件Ω在试验中必然发生. 相反地, 不可能事件在任何试验中不可能发生,必然事件与不可能事件可以说不是随机事件,但为了今后研究的方便, 我们还是把必然事件与不可能事件作为随机事件的两个极端情形来统一处理.

三、事件的运算

在一个样本空间中显然可以定义不止一个事件. 概率论的重要研究课题之一是希望从简单事件的概率推算出复杂的事件的概率. 在实际生活中, 往往要求我们同时考察几个在同样条件下的事件及它们之间的联系. 详细地分析事件之间的关系,不仅帮助我们更深刻地认识事件的本质, 而且可以大大简化一些复杂事件的概率计算.

下面就讨论事件间的关系及事件的运算. 先讨论两个事件A与B之间的关系.

若A中的每一个样本点都包含在B中,则记为$A\subset B$或$B\supset A$, 并称 **A 是 B 的特款**, 亦称**事件 B 包含了事件 A**, 这时事件A发生必然导致事件B发生. 例如若以A记"来到呼叫不超过 5 个", 以

B 记"来到呼叫不超过 6 个"，则 $A\subset B$. 显然对任何事件 A, 必有 $\Omega\supset A\supset\varnothing$.

如果 $A\supset B$ 与 $B\supset A$ 同时成立，则称 **A 与 B 等价**或称 **A 等于 B**, 记为 $A=B$, 等价的两个事件同时发生，因此可看作是一样的.

对于事件 A, 由所有不包含在 A 中的样本点所组成的事件称为 **A 的逆事件**, 或称 **A 的对立事件**, 记为 $\bar{A}$, $\bar{A}$ 表示 A 不发生. 例如若以 A 表示"来到呼叫不超过 5 个"，则 $\bar{A}$ 表示"来到呼叫超过 5 个". 显然，若 $\bar{A}$ 是 A 的对立事件，则 A 也是 $\bar{A}$ 的对立事件，即 $\bar{\bar{A}}=A$. 必然事件与不可能事件互为对立事件.

其次，对于事件 A 及事件 B, 定义两个新事件:

用 $A\cap B$ 或 AB 表示所有同时属于 A 及 B 的样本点的集合，称它为 **A 与 B 的交**, 事件 AB 表示事件 A 与事件 B 同时发生.

用 $A\cup B$ 表示至少属于 A 或 B 中的一个的所有样本点的集合，称它为 **A 与 B 的并**，事件 $A\cup B$ 表示事件 A 或事件 B 或它们二者发生，也即表示事件 A 与事件 B 至少发生一个.

若 $AB=\varnothing$, 则表示 A 与 B 不可能同时发生，称 **A 与 B 互不相容**. 样本点是互不相容的.

以后，对于互不相容事件 A 与 B，我们称它们的并为**和**，并记作 $A+B$.

用 $A-B$ 表示包含在 A 中而不包含在 B 中的样本点全体，称之为 **A 与 B 的差**，事件 $A-B$ 表示事件 A 发生而事件 B 不发生，显然 $A-B=A\bar{B}$.

在进行事件的运算时，关于它们的顺序作如下约定：先进行逆的运算，再进行交的运算，最后才进行并或差的运算.

用上面的记号可以把对立事件之间的关系表述如下: $A\cup\bar{A}=\Omega$, $A\cap\bar{A}=\varnothing$, 这也可以作为对立事件的定义. 显然 $\bar{A}=\Omega-A$.

我们用例 5 来说明这些关系. (*)中 30 种可能结果就是样本

点全体，它们构成必然事件 Ω.

A 记第一次摸得黑球，则它是由第 5 行及第 6 行的 10 个样本点构成；这时 $\bar{A}$ 表示第一次摸得白球，它由第一行至第四行的 20 个样本点构成. 显然 A 与 $\bar{A}$ 互不相容，而且 $A+\bar{A}=\Omega$.

B 记第二次摸得黑球，它由下列 10 个样本点构成：

$$(1,5),(1,6),(2,5),(2,6),(3,5)$$
$$(3,6),(4,5),(4,6),(5,6),(6,5)$$

事件 $A\cup B$ 表示第一次或第二次中至少有一次摸得黑球，它包含下列 18 个样本点：

$$(1,5),(1,6),(2,5),(2,6),(3,5),(3,6)$$
$$(4,5),(4,6),(5,1),(5,2),(5,3),(5,4)$$
$$(5,6),(6,1),(6,2),(6,3),(6,4),(6,5)$$

事件AB表示两次摸得黑球，它由下列两个样本点构成：(5, 6), (6, 5)，这是 A 与 B 共同包含的样本点. 因此 $C=AB$.

事件 $A-B$ 表示第一次摸得黑球而第二次摸出白球，包含了 (5, 1), (5, 2), (5, 3), (5, 4), (6, 1), (6, 2), (6, 3), (6, 4)等 8 个样本点.

有时用平面上某正方形中的图形来表示事件的关系较为直观，这种表示法称为文(Venn)图.

事件 $\Omega, A, \bar{A}, A\cup B, AB, A-B$ 在图 1 中分别以阴影表出. 不难理解，$B\supset A$ 相应于 A 的图形完全包含在 B 的图形中；A 和 B 互不相容，则相应于 A 和 B 的图形不相交.

不难把上面定义推广到多个事件的场合.

例如，对于 n 个事件 $A_1, A_2, \cdots, A_n$，用 $A_1\cup A_2\cup\cdots\cup A_n$ 或 $\bigcup_{i=1}^{n} A_i$ 表示 $A_1, A_2, \cdots, A_n$ 中至少发生一个，称为 $A_1, A_2, \cdots, A_n$ 的并，特别当 $A_1, A_2, \cdots, A_n$ 两两互不相容时，并称为和，并记作 A_1+

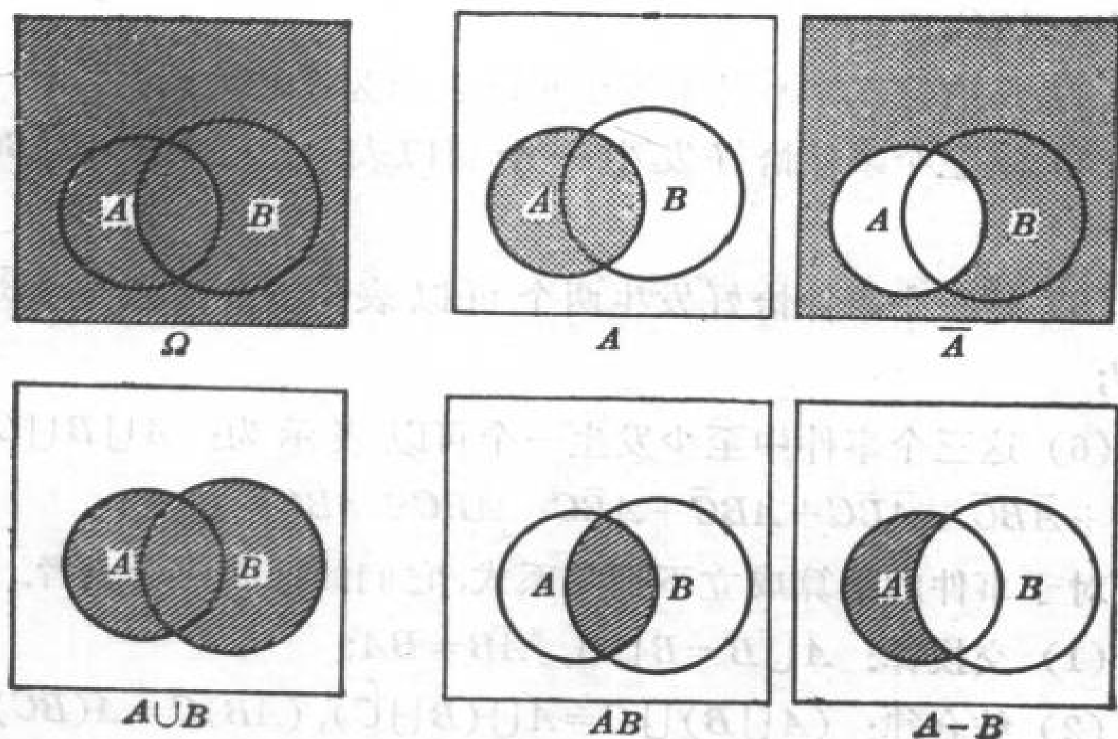


图1 事件运算

$A_2+\cdots+A_n$ 或 $\sum_{i=1}^{n}A_i$. 用 $A_1A_2\cdots A_n$ 或 $\bigcap_{i=1}^{n}A_i$ 表示 $A_1, A_2, \cdots, A_n$ 同时发生等等.

对于可列个事件的场合，我们定义

$$\bigcup_{i=1}^{\infty}A_i=\lim_{n\to\infty}\bigcup_{i=1}^{n}A_i,\quad \bigcap_{i=1}^{\infty}A_i=\lim_{n\to\infty}\bigcap_{i=1}^{n}A_i$$

熟悉集合论的读者或许早就发现，事件间的关系及运算与集合论中或布尔（Boole）代数中集合的关系及运算是完全相似的，而且这个相似在建立概率论的严格数学基础时非常重要. 不过，我们应该注意另一点，就是要学会用概率论的语言来解释这些关系及运算. 并且会用这些运算关系来表示一些事件.

［例6］ 若 A, B, C 是三个事件，则

（1） A 发生而 B 与 C 都不发生可以表示为: $A\bar{B}\bar{C}$ 或 $A-B-C$ 或 $A-(B\cup C)$;

（2） A 与 B 都发生而 C 不发生可以表示为: $AB\bar{C}$ 或 $AB-C$

或 $AB-ABC$;

(3) 所有这三个事件都发生可以表示为: ABC;

(4) 这三个事件恰好发生一个可以表示为: $A\bar{B}\bar{C}+\bar{A}B\bar{C}+\bar{A}\bar{B}C$;

(5) 这三个事件恰好发生两个可以表示为: $AB\bar{C}+A\bar{B}C+\bar{A}BC$;

(6) 这三个事件中至少发生一个可以表示为: $A\cup B\cup C$ 或 $A\bar{B}\bar{C}+\bar{A}B\bar{C}+\bar{A}\bar{B}C+AB\bar{C}+A\bar{B}C+\bar{A}BC+ABC$;

对于事件的运算成立下列关系式,它们的证明留给读者.

(1) 交换律: $A\cup B=B\cup A$, $AB=BA$;

(2) 结合律: $(A\cup B)\cup C=A\cup(B\cup C)$, $(AB)C=A(BC)$;

(3) 分配律: $(A\cup B)\cap C=AC\cup BC$,

$$(A\cap B)\cup C=(A\cup C)\cap(B\cup C)$$

(4) 德莫根(De Morgan)定理:

$$\overline{A_1\cup A_2}=\bar{A}_1\cap\bar{A}_2,\qquad \overline{A_1\cap A_2}=\bar{A}_1\cup\bar{A}_2$$

对于 n 个事件,甚至对于可列个事件,德莫根定理也成立.

四、有限样本空间

我们先考虑只有有限个样本点的样本空间,这种样本空间称为**有限样本空间**. 这是最简单的样本空间,研究它有助于深入研究更为复杂的样本空间.

若 Ω 是有限样本空间,其样本点为 $E_1, E_2, \cdots, E_n$, 在这种场合可以把 Ω 的任何子集都当作事件. 在这种样本空间中引进概率,只要对每个样本点 E_i 给定一个数与它对应,此数称为 E_i 的**概率**,并记之为 $P(E_i)$,它是非负的,而且满足

$$P(E_1)+P(E_2)+\cdots+P(E_n)=1$$

这样,我们对样本点定义了概率,用它来度量每个样本点出现的可能性的大小. 由此出发,我们不难定义更为一般的事件的概率.

定义 任何事件A的概率$P(A)$是A中各样本点的概率之和.

按照这个定义,显然有$P(\Omega)=1, 0\leqslant P(A)\leqslant 1$.

如在例5中,若定义每个样本点出现的概率均为$\frac{1}{30}$(这相当于假定各个球外形完全一样,并且摸球是随机的,各个球被摸到的机会均等),则得 $P(A)=\frac{10}{30}$, $P(B)=\frac{10}{30}$, $P(C)=\frac{18}{30}$, $P(A-B)=\frac{8}{30}$等等.

下节,我们将研究一种特殊的有限样本空间.

把上面做法推广到有可列个样本点的样本空间是不难的, 这种空间称为**离散样本空间**. 但是当把上面做法推广到不可列个样本点的场合,则会遇到实质性的困难, 对于这种一般场合的讨论,以后将逐渐展开.

§3. 古典概型

一、模型与计算公式

在讨论一般随机现象之前, 我们先讨论一类最简单的随机现象. 这种随机现象具有下列两个特征:

(1) 在观察或试验中它的全部可能结果只有有限个, 譬如为n个,记为$E_1, E_2, \cdots, E_n$,而且这些事件是两两互不相容的;

(2) 事件$E_1, E_2, \cdots, E_n$的发生或出现是等可能的,即它们发生的概率都一样.

这类随机现象在概率论发展初期即被注意, 许多最初的概率论结果也是对它作出的, 一般把这类随机现象的数学模型称为**古典概型**. 古典概型在概率论中占有相当重要的地位. 一方面, 由于它简单,对它的讨论有助于直观地理解概率论的许多基本概念,因此,我们常从讨论古典概型开始引入新的概念;另一方面, 古典

概型概率的计算在产品质量抽样检查等实际问题以及理论物理的研究中都有重要应用.

显然,古典概型是有限样本空间的一种特例. 可以选 $\Omega=\{E_1, E_2, \cdots, E_n\}$作为样本空间, 而且此时应有

$$P(E_1)=P(E_2)=\cdots=P(E_n)=\frac{1}{n}$$

对于任何事件 A, 它总可以表示为样本点之和,例如 $A=E_{i_1}+E_{i_2}+\cdots+E_{i_m}$, 因此由事件概率的定义

$$\begin{aligned}P(A)&=P(E_{i_1})+P(E_{i_2})+\cdots+P(E_{i_m})\\&=\frac{1}{n}+\frac{1}{n}+\cdots+\frac{1}{n}=\frac{m}{n}\end{aligned} \tag{1}$$

所以在古典概型中,事件 A 的概率是一个分数,其分母是样本点的总数 n, 而分子是事件 A 中所包含的样本点的个数 m, 由于 $E_{i_1}, E_{i_2}, \cdots, E_{i_m}$ 的出现必导致 A 的出现, 即它们的出现对 A 的出现"有利", 因此习惯上常称 $E_{i_1}, E_{i_2}, \cdots, E_{i_m}$ 是 A 的"有利场合", 这样,

$$P(A)=\frac{m}{n}=\frac{A\text{包含的样本点数}}{\text{样本点总数}}=\frac{A\text{的有利场合的数目}}{\text{样本点总数}} \tag{2}$$

法国数学家拉普拉斯(Laplace)在 1812 年把上式作为概率的一般定义. 现在通常称它为概率的古典定义, 因为它只适用于古典概型场合.

古典概型有着多方面应用,产品抽样检查就是其中之一.

产品抽样检查的技术,在各个生产部门中被广泛采用.许多大工厂产量很高,每天生产的产品数以万计,对这些产品的质量如果要进行全面的逐件检验经常是不可能的或是不经济的;另外,在有些情况下,产品的检验方法带有破坏性(如电灯泡寿命检验和棉纱强度试验),这样,最适宜的检验方法是采用抽样检查,即从产品中随机地抽出若干件来检验,根据检验结果来判断整批产品的质量.

关于产品的质量，可以有多种多样的衡量标准，例如可能要考虑产品的某种形状或尺寸，或把产品分成若干等级，我们先考虑最简单的情形，即把产品分成合格品(好品)与次品(废品)两个类型的场合。

假如产品的好坏从外形上看不出来，而且我们又是随机抽样，那么任何一件产品被抽到的可能性都一样，这正是古典概型。

有一个袋子，内装 a 只黑球，b 只白球，它们除颜色不同外，外形完全一样(以后若非特别声明，均作此假定)．这样一来，当我们从袋子中任意摸出一球时，这 $a+b$ 只球中的任意一只被摸到的可能性都一样．

若把黑球作为废品，白球作为好品，则这个摸球模型就可以描述产品抽样．假如产品分为更多等级，例如一等品，二等品，三等品，等外品等等，则可用装有多种颜色的摸球模型来描述．

这种模型化的方法能使问题更清楚，更容易看出其随机性本质而不致被个别情况下的具体属性所蒙蔽．不仅如此，这种抽象化的模型带有普遍性，它还可以描述许多别的具体问题，从而有着多方面应用．例如种水稻生产队的调查，某种疾病的抽查等都能用这个模型．

事实上，古典概型的大部分问题都能形象化地用摸球模型来描述．以后我们经常研究摸球模型，意义即在于此．

前节例 5 及其有关概率的计算是古典概型的一个例子，但并不是所有古典概型的事件的概率计算都这么容易．事实上，古典概型中许多概率的计算相当困难而富有技巧，计算的要点是给定样本点，并计算它的总数，而后再计算有利场合的数目．在这些计算中，经常要用到一些排列与组合公式．

二、基本的组合分析公式

1．全部组合分析公式的推导基于下列两条原理：

乘法原理 若进行 A_1 过程有 n_1 种方法，进行 A_2 有 n_2 种方法，则进行 A_1 过程后再接着进行 A_2 过程共有 $n_1\times n_2$ 种方法(图 1).

加法原理 若进行 A_1 过程有 n_1 种方法，进行 A_2 过程有 n_2 种方法，假定 A_1 过程与 A_2 过程是并行的，则进行过程 A_1 或过程 A_2 的方法共有 n_1+n_2 种(图 2).

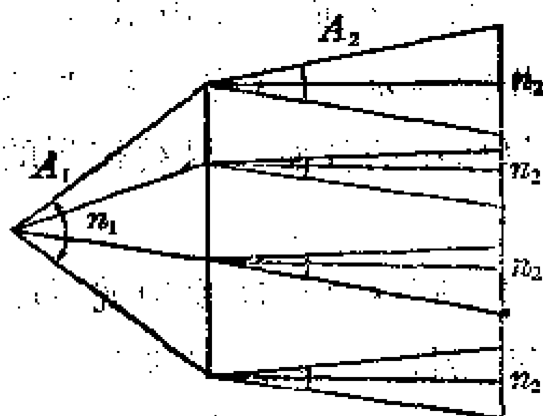


图 1

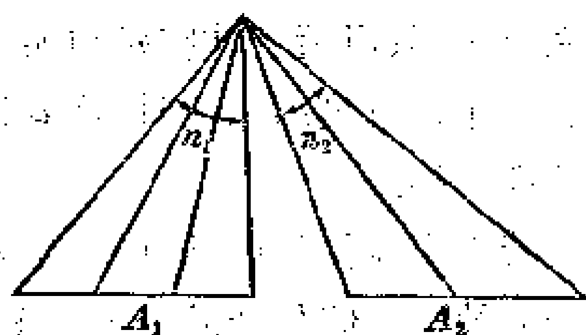


图 2

显然这二条原理可以拓广到多个过程的场合.

2. 排列:

从包含有 n 个元素的总体中取出 r 个来进行排列，这时既要考虑到取出的元素也要顾及其取出顺序.

这种排列可分为两类：第一种是有放回的选取，这时每次选取都是在全体元素中进行，同一元素可被重复选中；另一种是不放回选取，这时一个元素一旦被取出便立刻从总体中除去，因此每个元素至多被选中一次，在后一种情况，必有 $r\leqslant n$.

(1) 在有放回选取中，从 n 个元素中取出 r 个元素进行排列，这种排列称为**有重复的排列**，其总数共有 n^r 种.

(2) 在不放回选取中，从 n 个元素中取出 r 个元素进行排列，其总数为

$$A_n^r=n(n-1)(n-2)\cdots(n-r+1)$$

这种排列称为**选排列**. 特别当 $r=n$ 时，称为**全排列**.

(3) n 个元素的全排列数为 $P_n=n(n-1)\cdots 3\cdot 2\cdot 1=n!$

3. 组合:

(1) 从 n 个元素中取出 r 个元素而不考虑其顺序，称为**组合**，其总数为

$$C_n^r=\binom{n}{r}=\frac{A_n^r}{r!}=\frac{n(n-1)\cdots(n-r+1)}{r!}=\frac{n!}{r!(n-r)!}$$

这里$\binom{n}{r}$是二项展开式的系数，$(a+b)^n=\sum_{r=0}^{n}\binom{n}{r}a^r b^{n-r}$.

(2) 若 $r_1+r_2+\cdots+r_k=n$，把 n 个不同的元素分成 k 个部分，第一部分 r_1 个，第二部分 r_2 个，……，第 k 部分 r_k 个，则不同的分法有

$$\frac{n!}{r_1!r_2!\cdots r_k!} \tag{3}$$

种，上式中的数称为**多项系数**．因为它是$(x_1+x_2+\cdots+x_k)^n$ 展开式中 $x_1^{r_1}x_2^{r_2}\cdots x_k^{r_k}$ 的系数，当 $k=2$ 时，即为组合数．

(3) 若 n 个元素中有 n_1 个带足标"1"，n_2 个带足标"2"，……，n_k 个带足标"k"，且 $n_1+n_2+\cdots+n_k=n$，从这 n 个元素中取出 r 个，使得带有足标"i"的元素有 r_i 个$(1\leqslant i\leqslant k)$，而 $r_1+r_2+\cdots+r_k=r$，这时不同取法的总数为

$$\binom{n_1}{r_1}\binom{n_2}{r_2}\cdots\binom{n_k}{r_k} \tag{4}$$

这里当然要求 $r_i\leqslant n_i$.

4. 一些常用等式:

把排列公式推广到 r 是正整数而 n 是任意实数 x 的场合，有时是需要的，这时记

$$A_x^r=x(x-1)(x-2)\cdots(x-r+1)$$

同样定义

$$\binom{x}{r}=\frac{A_x^r}{r!}=\frac{x(x-1)(x-2)\cdots(x-r+1)}{r!}$$

及 $0!=1$，$\binom{x}{0}=1$.

对于正整数 n，若 $r>n$，则$\binom{n}{r}=0$.

这样一来二项系数有性质

$$\binom{n}{k}=\binom{n}{n-k}$$

$$\binom{-a}{k}=(-1)^k\binom{a+k-1}{k}$$

由于

$$(1+x)^n=\sum_{r=0}^{n}\binom{n}{r}x^r$$

故

$$\binom{n}{0}+\binom{n}{1}+\binom{n}{2}+\cdots+\binom{n}{n}=2^n$$

利用幂级数乘法又可以证明

$$\binom{a}{0}\binom{b}{n}+\binom{a}{1}\binom{b}{n-1}+\cdots+\binom{a}{n}\binom{b}{0}=\binom{a+b}{n} \tag{5}$$

特别地

$$\binom{n}{0}\binom{n}{n}+\binom{n}{1}\binom{n}{n-1}+\cdots+\binom{n}{n}\binom{n}{0}=\binom{2n}{n}$$

即

$$\binom{n}{0}^2+\binom{n}{1}^2+\cdots+\binom{n}{n}^2=\binom{2n}{n}$$

三、概率直接计算的例子

[例 1] 一部四本头的文集按任意次序放到书架上去，问各册自右向左或自左向右恰成 1, 2, 3, 4 的顺序的概率是多少？

[解] 若以 a, b, c, d 分别表示自左向右排列的书的卷号，则上述文集放置的方式可与向量(a, b, c, d)建立一一对应，因为 a,

b, c, d 取值于 1, 2, 3, 4, 因此这种向量的总数相当于 4 个元素的全排列数 $4!=24$, 由于文集按"任意的"次序放到书架上去, 因此这 24 种排列中出现任意一种的可能性都相同, 这是古典概型概率, 其有利场合有 2 种, 即自左向右或自右向左成 1, 2, 3, 4 顺序, 因此所求概率为$\frac{2}{24}=\frac{1}{12}$.

[例 2] 有 10 个电阻, 其电阻值分别为 1 Ω, 2 Ω, …, 10 Ω, 从中取出三个, 要求取出的三个电阻, 一个小于 5 Ω, 一个等于 5 Ω, 另一个大于 5 Ω, 问取一次就能达到要求的概率.

[解] 把从 10 个电阻中取出 3 个的各种可能取法作为样本点全体, 这是古典概型, 其总数为$\binom{10}{3}$, 有利场合为$\binom{4}{1}\binom{1}{1}\binom{5}{1}$, 故所求概率为

$$P=\frac{\binom{4}{1}\binom{1}{1}\binom{5}{1}}{\binom{10}{3}}=\frac{1}{6}$$

[例 3] 某城有 N 部卡车, 车牌号从 1 到 N, 有一个外地人到该城去, 把遇到的 n 部车子的牌号抄下(可能重复抄到某些车牌号), 问抄到的最大号码正好为 k 的概率. $(1\leqslant k\leqslant N)$

[解] 这种抄法可以看作是对 N 个车牌号进行 n 次有放回的抽样. 所有可能的抽法共有 N^n 种, 以它为样本点全体. 由于每部卡车被遇到的机会可以认为相同, 因此这是一个古典概型概率的计算问题, 有利场合数可以这样考虑: 先考虑最大车牌号不大于 k 的取法, 这样取法共有 k^n 种, 再考虑最大车牌号不大于 $k-1$ 的取法, 其数目有 $(k-1)^n$ 种, 因此有 $k^n-(k-1)^n$ 种取法其最大车牌号正好为 k, 这就是有利场合的数目, 因而所求概率为

$$P=\frac{k^n-(k-1)^n}{N^n}$$

[例 4] 设有 n 个球，每个都能以同样的概率 $\frac{1}{N}$ 落到 N 个格子 $(N\geqslant n)$ 的每一个格子中，试求：

(1) 某指定的 n 个格子中各有一个球的概率；

(2) 任何 n 个格子中各有一个球的概率.

[解] 这是一个古典概型问题，由于每个球可落入 N 个格子中的任一个，所以 n 个球在 N 个格子中的分布相当于从 N 个元素中选取 n 个进行有重复的排列，故共有 N^n 种可能分布.

在第一个问题中，有利场合相当于 n 个球在那指定的 n 个格子中全排列，总数为 $n!$，因而所求概率为

$$P_1=\frac{n!}{N^n}$$

在第二个问题中，n 个房间可以任意，即可以从 N 个房间中任意选出 n 个来，这种选法共有 $\binom{N}{n}$ 种，对于每种选定的 n 个房间，有利场合正如第一个问题一样为 $n!$，故所求概率为

$$P_2=\frac{\binom{N}{n}n!}{N^n}=\frac{N!}{N^n(N-n)!}$$

这个例子是古典概型中一个很典型的问题，不少实际问题可以归结为它.

例如，若把球解释为粒子，把格子解释为相空间中的小区域，则这个问题便相应于统计物理学中的马克斯威尔-波尔茨曼(Maxwell-Boltzmann)统计.

概率论历史上有一个颇为有名的问题：要求参加某次集会的 n 个人中没有两个人生日相同的概率. 若把 n 个人看作上面问题中的 n 个球，而把一年的 365 天作为格子，则 $N=365$，这时 P_2 就给出所求的概率. 例如当 $n=40$ 时，$P_2=0.109$，这个概率是意外的小.

［例 5］ 袋中有 a 只黑球，b 只白球，它们除颜色不同外，其他方面没有差别，现在把球随机地一只只摸出来，求第 k 次摸出的一只球是黑球的概率（$1\leqslant k\leqslant a+b$）。

［第一种解法］ 把 a 只黑球及 b 只白球都看作是不同的（例如设想把它们进行编号），若把摸出的球依次放在排列成一直线的 $a+b$ 个位置上，则可能的排列法相当于把 $a+b$ 个元素进行全排列，总数为 $(a+b)!$，把它们作为样本点全体。有利场合数为 $a\times(a+b-1)!$，这是因为第 k 次摸得黑球有 a 种取法，而另外 $(a+b-1)$ 次摸球相当于 $a+b-1$ 只球进行全排列，有 $(a+b-1)!$ 种构成法，故所求概率为

$$P_k=\frac{a\times(a+b-1)!}{(a+b)!}=\frac{a}{a+b}$$

这个结果与 k 无关。回想一下，就会发觉这与我们平常的生活经验是一致的。例如在体育比赛中进行抽签，对各队机会均等，与抽签的先后次序无关。

［第二种解法］ 把 a 只黑球看作是没有区别的，把 b 只白球也看作是没有区别的。仍把摸出的球依次放在排列成一直线的 $a+b$ 个位置上，因若把 a 只黑球的位置固定下来则其他位置必然是放白球，而黑球的位置可以有 $\binom{a+b}{a}$ 种放法，以这种放法作为样本点。这时有利场合数为 $\binom{a+b-1}{a-1}$，这是由于第 k 次摸得黑球，这个位置必须放黑球，剩下的黑球可以在 $a+b-1$ 个位置上任取 $a-1$ 个位置，因此共有 $\binom{a+b-1}{a-1}$ 种放法。所以所求概率为

$$P_k=\frac{\binom{a+b-1}{a-1}}{\binom{a+b}{a}}=\frac{a}{a+b}$$

两种不同的解法答案相同．注意考察一下两种解法的不同，就会发现主要在于选取的样本空间不同．在前一种解法中把球看作是“有个性的”，而在后一种解法中则对同色球不加区别，因此在第一种解法中要顾及各黑球及各白球间的顺序而用排列，第二种解法则不注意顺序而用组合，但最后还是得出了相同的答案．

这种情况的产生并不奇怪，这说明对于同一随机现象，可以用不同的模型来描述，只要方法正确，结论总是一致的．在这个例子中，第二种解法中的每一个样本点是由第一种解法中的 $a!\cdot b!$ 个样本点合并而成的．

这个例子告诉我们，在计算样本点总数及有利场合数时，必须对同一个确定的样本空间考虑，因此其中一个考虑顺序，另一个也必须考虑顺序，否则结果一定不正确．

既然同一个随机现象可用不同的样本空间来描述，因此对同一个概率也常常有多种不同的求法，我们应逐步训练自己能采用最简便的方法解题，为此熟悉同一问题的多种不同解法是重要的．

例如，对例5就存在着多种不同的解法，上面提供的只是比较自然的两种．注意到在这两种解法中，我们对不同的 k 用的是同一个样本空间，也就是说，我们构造了一个可以描述 $a+b$ 次摸球的样本空间，并利用它一举解决了“第 $k(1\leqslant k\leqslant a+b)$ 次摸得黑球”这一概率的计算．假如允许对不同的 k 用不同的样本空间，则我们完全可以构造一个只包含前 k 次试验，甚至只包含第 k 次试验的样本空间，这时也能求得有关概率．特别是选用最后一种样本空间简直马上可以看出正确答案，不过这种做法对初学者或许不那么容易理解．

四、二项分布与超几何分布

产品抽样检验有两类，即有放回抽样与不放回抽样．在有放回抽样中，被抽出的产品检验后仍放回产品堆中，再抽第二次．因

此这件产品以后仍然可能再次被抽到. 更常用的是第二种方法,即不放回的抽样方法,这时被抽到的产品不再放回,因而以后不会再被抽到. 与此相应的,我们的摸球模型也假定为有放回与不放回摸球两种,这两个情形得到的结果是不同的.

下面是古典概型概率计算中一个典型问题,它有着多方面应用,特别在产品检验方面起很大作用.

[例6] 如果某批产品中有 a 件次品 b 件好品,我们采用有放回及不放回抽样方式从中抽 n 件产品,问正好有 k 件是次品的概率各是多少?

所求的概率显然与抽样方式有关,下面我们分别来讨论.

[有放回抽样场合] 把 $a+b$ 个产品进行编号,有放回抽 n 次,把可能的重复排列全体作为样本点,总数为 $(a+b)^n$,其中有利场合(即次品正好出现 k 次)的数目是 $\binom{n}{k}a^k b^{n-k}$,故所求概率为

$$b_k=\frac{\binom{n}{k}a^k b^{n-k}}{(a+b)^n}=\binom{n}{k}\left(\frac{a}{a+b}\right)^k\left(\frac{b}{a+b}\right)^{n-k} \tag{6}$$

b_k 是二项式 $\left(\frac{a}{a+b}+\frac{b}{a+b}\right)^n$ 展开式的一般项,**上述概率称为二项分布**(分布一词的意义将在第三章阐明). 关于二项分布更一般的讨论在以后各章陆续进行.

[不放回抽样场合] 从 $a+b$ 个产品中取出 n 个产品的可能组合全体作为样本点,总数为 $\binom{a+b}{n}$,有利场合数为 $\binom{a}{k}\binom{b}{n-k}$,故所求概率为

$$h_k=\frac{\binom{a}{k}\binom{b}{n-k}}{\binom{a+b}{n}} \tag{7}$$

这个概率称为超几何分布.

从直观上看，当产品总数很大而抽样数不大时，采用有放回抽样与采用不放回抽样，影响应该不大.

事实上，因为

$$h_k=\frac{\binom{a}{k}\binom{b}{n-k}}{\binom{a+b}{n}}=\frac{\frac{A_a^k}{k!}\cdot\frac{A_b^{n-k}}{(n-k)!}}{\frac{A_{a+b}^n}{n!}}=\binom{n}{k}\frac{A_a^k A_b^{n-k}}{A_{a+b}^n}$$

$$=\binom{n}{k}\frac{a^k b^{n-k}}{(a+b)^n}\cdot\frac{\frac{A_a^k}{a^k}\cdot\frac{A_b^{n-k}}{b^{n-k}}}{\frac{A_{a+b}^n}{(a+b)^n}}$$

而当 k 比 a 小得多，$n-k$ 比 b 小得多时：

$$\frac{\frac{A_a^k}{a^k}\cdot\frac{A_b^{n-k}}{b^{n-k}}}{\frac{A_{a+b}^n}{(a+b)^n}}\approx 1$$

因此

$$h_k\approx b_k$$

在实际工作中，抽样一般都采用不放回方式，因此计算废品数为 k 的概率时应该用超几何分布，(7)的数值计算较繁复. 不过，若产品数甚大而抽样数不太大，则可利用上述性质，计算二项分布作为近似值，这时有许多专门表格可查，这样可以大大节省计算工作量.

利用上例结果，马上可以计算下列概率：若一批产品共有 N 件，其中有次品 $M(M<N)$ 件，今抽取 n 件，则其中恰有 m 件次品的概率是

$$P_m=\frac{\binom{M}{m}\binom{N-M}{n-m}}{\binom{N}{n}},\qquad \begin{array}{l}0\leqslant n\leqslant N,\ 0\leqslant m\leqslant M\\ 0\leqslant n-m\leqslant N-M\end{array}$$

这是超几何分布的另一种常见形式.

学到这里，细心的读者可能会发觉这样一个矛盾，在我们前面讨论中都假定产品中的次品数已知，然后根据它来计算种种概率，而在实际问题中，情况恰恰相反，次品数是未知的，并且正是我们希望通过抽样检验来确定的.

这个矛盾可通过下面办法来解决.

不难理解，抽出来的样本的质量情况在某种程度上反映了整批产品的质量情况，例如，如果整批产品中次品很多，则抽查的样本中含有次品的可能性就相当大；反之，若产品中极少次品，则从中抽查一、两只样本而得到次品的可能性就很小，因而样本中所含次品数的多少就为我们估计整批产品中的次品数提供了某种根据．例如为了确定某批产品的次品率，通常采用的方法是从这批产品中抽若干个产品作为样本来检验，并用样本的次品率来估计整批产品的次品率．关于这个课题的研究，构成了数理统计的重要内容.

由于抽样带有随机性，因而不同的抽样可能得到不同的结果，所以我们有必要对各种结果出现的可能性大小进行讨论，这为我们根据样本情况推断整批产品情况提供了理论依据，这种研究是概率论的任务．从这里也看出，概率论与数理统计有着很密切的联系.

[例 7]　从某鱼池中捕得 1200 条鱼，做了红色的记号之后再放回池中，经过适当的时间后，再从池中捕 1000 条鱼，计算其中有红色记号的鱼的数目，共有 100 条，试估计鱼池中共有多少条鱼.

[解]　设池中共有 n 条鱼，n 未知，是我们要估计的．更一般地，设第一次捕得的鱼有 n_1 条(在我们的场合 $n_1=1200$)，第二次捕得 r 条(在我们的场合 $r=1000$)，而其中有记号的有 k 条，(在我们的场合 $k=100$)。

现在，在第二次捕鱼中有 k 条有记号鱼的概率由(7)给出：

$$p_k(n)=\binom{n_1}{k}\binom{n-n_1}{r-k}\bigg/\binom{n}{r}$$

因为实际上在 r 条鱼中有 k 条有记号，因此我们求 n 使得上式概率达到最大，并把这个数值作为池中鱼数的估计.

由于

$$\frac{p_k(n)}{p_k(n-1)}=\frac{(n-n_1)(n-r)}{(n-n_1-r+k)n}=\frac{n^2-nn_1-nr+n_1r}{n^2-nn_1-nr+nk}=\rho$$

因此当 $nk<n_1r$ 时，$\rho>1$，而当 $nk>n_1r$ 时，$\rho<1$，即 $p_k(n)$ 当 $n<\frac{n_1r}{k}$ 时是 n 的增函数，而当 $n>\frac{n_1r}{k}$ 时是 n 的减函数，所以当 n 等于

$$\hat{n}=\left[\frac{n_1r}{k}\right]$$

时，$p_k(n)$ 达到最大值. 这样我们把 $\left[\frac{n_1r}{k}\right]$ 作为鱼池中鱼总数 n 的估计量.

在这个例子中：

$$\hat{n}=\frac{1200\times1000}{100}=12000$$

五、概率的基本性质

加法定理 两个互不相容事件 A 与 B 的和的概率等于事件 A 与 B 的概率之和，即

$$P(A+B)=P(A)+P(B) \tag{8}$$

[证明] 若样本空间为 $\Omega=\{E_1, E_2, \cdots, E_n\}$，设 A 包含样本点 $E_{i_1}, E_{i_2}, \cdots, E_{i_k}, P(A)=\frac{k}{n}$. 再设 B 包含样本点 $E_{j_1}, E_{j_2}, \cdots, E_{j_l}$，因此 $P(B)=\frac{l}{n}$. 由于 A 和 B 互不相容，所以 $E_{i_1}, E_{i_2}, \cdots, E_{i_k}, E_{j_1}, E_{j_2}, \cdots, E_{j_l}$ 都是不同的，因此 $A+B$ 中含有 $k+l$ 个样本点，

按照概率计算公式(1),

$$P(A+B)=\frac{k+l}{n}=\frac{k}{n}+\frac{l}{n}=P(A)+P(B)$$

从而证明了加法定理.

不难把上面结果推广到有限个事件的场合，即若 A_1, A_2, $\cdots$, A_m 是 m 个两两互不相容的事件，则

$$P(A_1+A_2+\cdots+A_m)=P(A_1)+P(A_2)+\cdots+P(A_m) \quad (9)$$

这个结论可籍助(8)用归纳法证明.

系 $P(\bar{A})=1-P(A)$. (10)

[证] 由于 $A+\bar{A}=\Omega$, $A\bar{A}=\varnothing$ 所以

$$P(\Omega)=P(A+\bar{A})=P(A)+P(\bar{A})=1$$

因此

$$P(\bar{A})=1-P(A)$$

关系式(10)很简单，意义也很明确，但相当有用，它为计算某些事件的概率提供了很大方便.

[例 8] 一袋中装有 $N-1$ 只黑球及 1 只白球，每次从袋中随机地摸出一球，并换入一只黑球，这样继续下去，问第 k 次摸球时摸到黑球的概率是多少?

[解] 若以 A 表示第 k 次摸到黑球这一事件，则 $\bar{A}$ 表示第 k 次摸到白球. 现在计算 $P(\bar{A})$.

因为袋中只有一只白球，而每次摸出白球总是换入黑球，故为了在第 k 次摸到白球，则前面的 $k-1$ 次一定不能摸到白球. 因此 $\bar{A}$ 等价于下列事件，在前 $k-1$ 次摸球时都摸出黑球而第 k 次摸出白球，这一事件的概率为

$$\frac{(N-1)^{k-1}\cdot 1}{N^k}=\left(1-\frac{1}{N}\right)^{k-1}\cdot\frac{1}{N}$$

这样

$$P(A)=1-P(\bar{A})=1-\left(1-\frac{1}{N}\right)^{k-1}\cdot\frac{1}{N}$$

公式(10)的应用使这个题目很快获得解决，假如直接计算$P(A)$则困难得多. 从公式(10)中容易看出，只要$P(A)$或$P(\bar{A})$中的任何一个知道了，则可以求得另一个. 在不同问题中，有的求$P(A)$容易，求$P(\bar{A})$困难；有的正好相反. 利用(10)，我们就可以先求容易的一个，再去求另一个.

从古典概型的概率研究中，我们发现概率有下面三个基本性质：

(i) 对于任何事件A，$P(A)\geqslant 0$； (11)

(ii) $P(\Omega)=1$； (12)

(iii) 若$A_1, A_2, \cdots, A_m$两两互不相容，则

$$P(A_1+A_2+\cdots+A_m)=P(A_1)+P(A_2)+\cdots+P(A_m) \quad (13)$$

第一个性质称为概率的**非负性**，第二个性质称为概率的**规范性**，第三个性质称为概率的**(有限)可加性**.

§4. 几何概率

一、例子与计算公式

在古典概型中利用等可能性的概念，成功地计算了某一类问题的概率；不过，古典概型要求可能场合的总数必须有限. 因此历史上有不少人企图把这种做法推广到有无限多结果而又有某种等可能性的场合. 这类问题一般可以通过几何方法来求解.

先从几个简单的例子开始.

[例1] 某人午觉醒来，发觉表停了，他打开收音机，想听电台报时，求他等待的时间短于10分钟的概率.

[例2] 如果在一个5万平方公里的海域里有表面积达40平方公里的大陆架贮藏着石油，假如在这海域里随意选定一点钻

探，问钻到石油的概率是多少？

[例 3] 在 400 毫升自来水中有一个大肠杆菌，今从中随机取出 2 毫升水样放到显微镜下观察，求发现大肠杆菌的概率．

一种相当自然的答案是认为例 1 所求的概率等于$\frac{1}{6}$，例 2 中钻到石油的概率等于$\frac{8}{10000}$，而例 3 所求的概率等于$\frac{1}{200}$．在求出这些概率时，我们事实上是利用了几何的方法，并假定了某种等可能性．

在例 1 中，因为电台每小时报时一次，我们自然认为这个人打开收音机时处于两次报时之间，例如(13:00, 14:00)，而且取各点的可能性一样，要遇到等待时间短于 10 分钟，只有当他打开收音机的时间正好处于 13:50 至 14:00 之间才有可能，相应的概率是$\frac{10}{60}=\frac{1}{6}$．

在例 2 中，由于选点的随机性，可以认为该海域中各点被选中的可能性是一样的，因而所求概率自然认为等于贮油海域的面积与整个海域面积之比，即等于$\frac{40}{50000}$．

同样地，例 3 中由于取水样的随机性，所求概率等于水样的体积与总体积之比$\frac{2}{400}$．

总之，在这类问题中，试验的可能结果是某区域Ω中的一个点．这个区域可以是一维的，也可以是二维的，还可以是三维的，甚至可以是 n 维的，这时不管是可能结果全体或是我们所感兴趣的结果都是无限的．因而等可能性是通过下列方式来赋予意义的：落在某区域 g 的概率与区域 g 的测度(长度，面积，体积等等)成正比并且与其位置及形状无关．

因此，若以 A_g 记“在区域Ω中随机地取一点，而该点落在区域

g 中"这一事件，则其概率定义为

$$P(A_g)=\frac{g\text{ 的测度}}{\Omega\text{ 的测度}} \tag{1}$$

据此定义，则上述诸例之解是明显的．下面再举一个例子．

[例 4] （会面问题）两人相约 7 点到 8 点在某地会面，先到者等候另一人 20 分钟，过时就可离去，试求这两人能会面的概率．

[解] 以 x，y 分别表示两人到达时刻，则会面的充要条件为

$$|x-y|\leqslant 20$$

这是一个几何概率问题，可能的结果全体是边长为 60 的正方形里的点，能会面的点的区域用阴影标出（图 1）．所求概率为

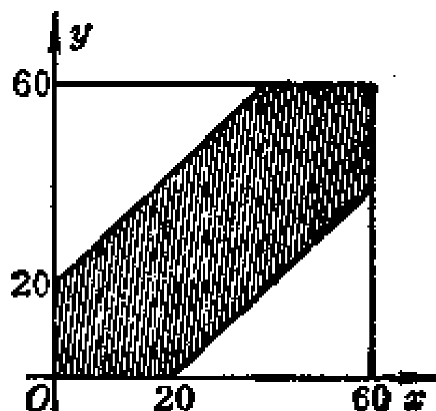


图 1

$$p=\frac{60^2-40^2}{60^2}=\frac{5}{9}$$

二、蒲丰(Buffon)问题

1777 年法国科学家蒲丰提出了下列著名问题，这是几何概率的一个早期例子．

[投针问题] 平面上画着一些平行线，它们之间的距离都等于 a，向此平面任投一长度为 l $(l<a)$ 的针，试求此针与任一平行线相交的概率．

[解] 以 x 表示针的中点到最近的一条平行线的距离，φ 表示针与平行线的交角．针与平行线的位置关系见图 2．

显然有 $0\leqslant x\leqslant\frac{a}{2}$，$0\leqslant\varphi\leqslant\pi$，以 G 表示边长为 $\frac{a}{2}$ 及 π 的长方形．为使针与平行线相交，必须 $x\leqslant\frac{l}{2}\sin\varphi$，满足这个关系式的区域记为 g，在图 2 中用阴影表出，所求的概率为

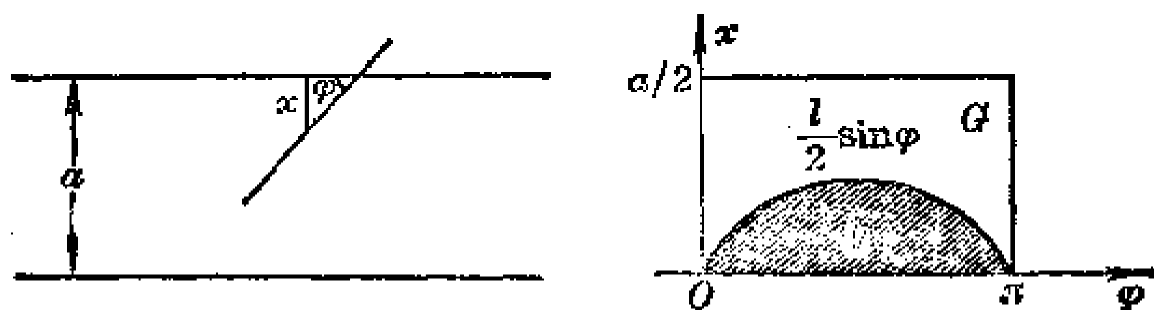


图 2

$$p=\frac{g\text{的面积}}{G\text{的面积}}=\frac{\frac{1}{2}\int_0^\pi l\sin\varphi d\varphi}{\frac{1}{2}a\pi}=\frac{2l}{\pi a} \tag{2}$$

由于最后的答案与π有关，因此不少人想利用它来计算π的数值，其方法是投针N次，计算针与线相交的次数n，再以频率值$\frac{n}{N}$作为概率p之值代入(2)，求得

$$\pi=\frac{2lN}{an} \tag{3}$$

下表给出了这些试验的有关资料(把a折算为1)

表1 投针试验的历史资料[①]

实验者	年份	针长	投掷次数	相交次数	π的实验值
Wolf	1850	0.8	5000	2532	3.1596
Smith	1855	0.6	3204	1218.5	3.1554
De Morgan, C.	1860	1.0	600	382.5	3.137
Fox	1884	0.75	1030	489	3.1595
Lazzerini	1901	0.83	3408	1808	3.1415929
Reina	1925	0.5419	2520	859	3.1795

值得注意的是这里采用的方法：建立一个概率模型，它与某些我们感兴趣的量——这里是常数π——有关，然后设计适当的随机试验，并通过这个试验的结果来确定这些量.

① 引自 Gridgeman, N. T. Geometric probability and the number π. Scripta Mathematica, 25(1960), 183—195.

现在，随着电子数字计算机的发展，已按照上述思路建立起一类新的方法，称蒙特卡洛(Monte-Carlo)方法。

三、贝特朗(Bertrand)奇论

几何概率在现代概率概念的发展中曾经起过重大作用。十九世纪时，不少人相信，只要找到适当的等可能性描述，就可以给概率问题以唯一的解答，然而有人却构造出这样的例子，它包含着几种似乎都同样有理但却互相矛盾的答案，下面就是一个著名的例子。

[贝特朗奇论] 在半径为1的圆内随机地取一条弦，问其长超过该圆内接等边三角形的边长$\sqrt{3}$的概率等于多少？

这是一个几何概率问题，但是基于对术语"随机地"的含义的不同解释，这个问题却存在多种不同答案，下面是其中的三种。

[解法一] 任何弦交圆周二点，不失一般性，先固定其中一点于圆周上，以此点为顶点作一等边三角形，显然只有落入此三角形内的弦才满足要求，这种弦的另一端跑过的弧长为整个圆周的$\frac{1}{3}$，故所求概率等于$\frac{1}{3}$(见图3)。

[解法二] 弦长只跟它与圆心的距离有关，而与方向无关，因此可以假定它垂直于某一直径，当且仅当它与

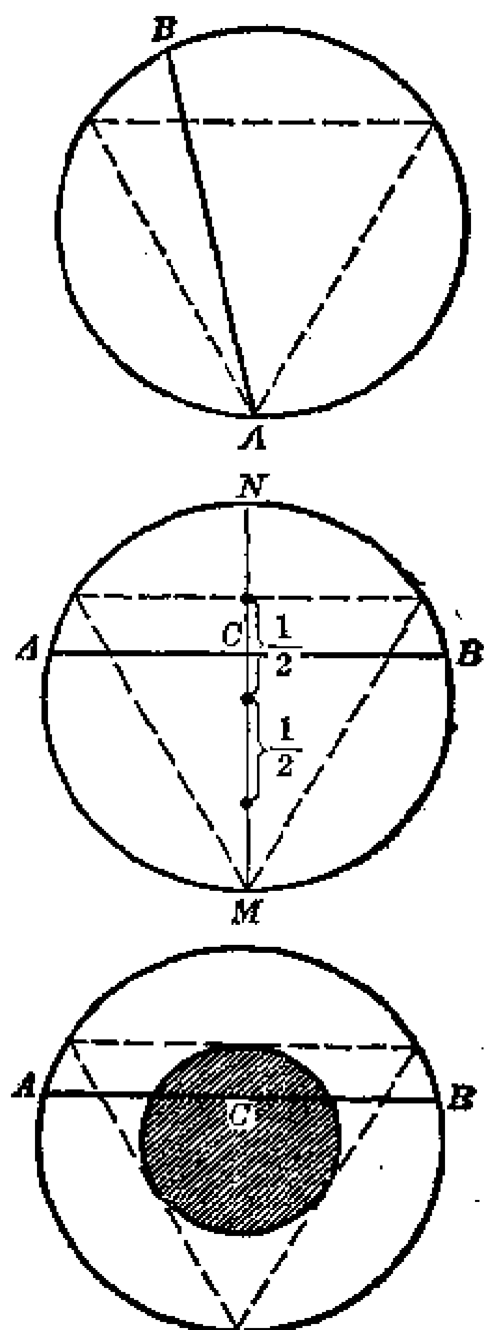


图 3

圆心的距离小于$\frac{1}{2}$时，其长才大于$\sqrt{3}$，因此所求概率为$\frac{1}{2}$(见图 3)．

［解法三］ 弦被其中点唯一确定，当且仅当其中点属于半径为$\frac{1}{2}$的同心圆内时，弦长大于$\sqrt{3}$，此小圆面积为大圆面积的$\frac{1}{4}$，故所求概率等于$\frac{1}{4}$(见图 3)．

同一问题有三种不同的答案，细究其原因，发现是在取弦时采用不同的等可能性假定．在第一种解法中，假定端点在圆周上均匀分布，在第二种解法中则假定弦的中点在直径上均匀分布，而在第三种解法中又假定弦的中点在圆内均匀分布．这三种答案是针对三种不同的随机试验，对于各自的随机试验而言，它们都是正确的．

因此在使用术语“随机”、“等可能”、“均匀分布”等时，应明确指明其含义；这又因试验而异．

由于采用等可能性来定义概率有这种困难，因此后来就选择另外的途径．即在定义概率这一基本概念时只指明概率应具有的基本性质，而把具体概率的给定放在一边．这样做的好处是能针对不同的随机试验给定适当的概率．

与概率的频率解释及古典概型一样，几何概率的研究对于我们了解应要求概率具有哪些基本性质是很有帮助的．

四、几何概率基本性质

几何概率的定义及计算与几何图形的测度密切相关，因此所考虑的事件应是某种可定义测度的集合．这类集合的并、交也还应该是事件，甚至对它们的可列次并、交也应有这个要求．例如考察在$[0,1)$中投一个点的随机试验，若以A记该点落入$\left(0,\frac{1}{2}\right)$中这

个事件，而以A_n记该点落入$\left[\frac{1}{2^{n+1}},\frac{1}{2^n}\right)$中这一事件，$n=1,2,\cdots$，则

$$A=\sum_{n=1}^{\infty}A_n \tag{4}$$

如果假定所投的点落入某区间的概率等于该区间的长度，则$P(A)=\frac{1}{2}$，而$P(A_n)=\frac{1}{2^{n+1}}$，这时有

$$P(A)=\sum_{n=1}^{\infty}P(A_n) \tag{5}$$

这里我们碰到了事件及概率的可列运算.

综上所述，几何概率应具有如下性质：

(i) 对任何事件A，$P(A)\geqslant 0$；

(ii) $P(\Omega)=1$；

(iii) 若$A_1,A_2,\cdots$两两互不相容，则

$$P\left(\sum_{n=1}^{\infty}A_n\right)=\sum_{n=1}^{\infty}P(A_n) \tag{6}$$

前两个性质与古典型概率相同，(6)则要求对可列个两两互不相容事件成立，这性质称为**可列可加性**.

§5. 概 率 空 间

一、走向概率论公理化结构

到本世纪初，概率论的各个领域已经得到了大量的成果，而且人们对概率论在其他基础学科和工程技术上的应用也出现了越来越大的兴趣. 但是直到那时为止，关于概率论的一些基本概念——例如事件、概率等等——却没有明确的定义，这是一个很大的矛盾，这个矛盾不仅可能导致贝特朗奇论那样怪现象产生，而且也使许多人对概率的客观含义甚至概率论结论的可应用性都产生了怀

疑. 因此可以说到那时为止，概率论作为一个数学分支来说，还缺乏严格的理论基础，这就大大妨碍了它的进一步发展.

在概率论发展早期，所研究的随机现象比较简单，大部分可以归入古典概型. 对这种模型，概率的计算可以通过某种等可能性的假设进行，其结果也有相当明确的解释. 这种成功使得人们试图通过给定某种等可能性来定义概率. 于是，由拉普拉斯给出的概率的古典定义在整个 19 世纪被人们广泛接受. 但是，这种定义的局限性很快也暴露了出来，它既要求试验的可能结果总数有限，又要求某种等可能性，所以它的适用范围有限. 当把这个结果推广到有无限多种可能结果的场合，例如几何概率时，不但适用范围仍然有限，而且还出现了新问题. 总之，对一般的随机现象明确地定义概率及其它基本概念，在那时成了一个突出的问题.

解决这个问题的时机也在不断成熟. 首先是通过对概率论的两个最基本概念——事件与概率的长期研究，发觉事件的运算与集合的运算完全相似，概率与测度有相同性质，这个事实随着当时在实变函数论中关于勒贝格 (Lebesgue) 测度和积分的研究以及一般抽象测度和积分理论的发展而日益明确起来.

另外，十九世纪末以来，数学的各个分支广泛流行着一股公理化潮流，这个流派主张把最基本的假定公理化，其他结论则由它们经过演绎导出.

在这种背景下，1933 年，苏联数学家柯尔莫哥洛夫(A. H. Колмогоров) 提出了概率论公理化结构，这个结构综合了前人成果，明确定义了基本概念，使概率论成为严谨的数学分支，对近几十年来概率论的迅速发展起了积极作用. 柯尔莫哥洛夫的这个理论已被广泛接受，本书的各部分就是在这个结构下展开的.

二、事件域

在公理化结构中，事件不是最基本的概念，它通过更基本的

概念——样本点来定义．**样本点**相应于随机试验的结果，我们已在前三节中进行过不少描述．在古典概型中，它们是可能结果全体，被用来定义一般的事件．在几何概率中，它们相应于区域Ω中的点，也被用来定义一般的事件．以后，我们把样本点ω看作抽象的点，它们的全体构成**样本空间**Ω.

正如§2中所做的那样，我们把事件A定义为Ω的一个子集，它包含若干样本点，事件A发生当且仅当A所包含的样本点中有一个发生．

一般并不把Ω的一切子集都作为事件，因为这将对给定概率带来困难，譬如在几何概率中，若把不可测集也作为事件，将带来不可克服的困难．

另一方面，又必须把问题中感兴趣的事件都包括进来．例如若A是事件，则应要求$\bar{A}$也是事件；若A与B是事件，则$A\cup B$及AB也应是事件．当样本空间Ω由无限多个点构成时——在几何概率中就是如此——显然还得考虑可列个事件的并与交．此外，把Ω及$\varnothing$作为事件有很大方便．

总之，我们若把事件的全体记为$\mathscr{F}$，它是由Ω的一些子集构成的集类．而且为了使讨论便于进行，还得对$\mathscr{F}$加上某些限制：

(i) $\Omega\in\mathscr{F}$;

(ii) 若$A\in\mathscr{F}$，则$\bar{A}\in\mathscr{F}$;

(iii) 若$A_n\in\mathscr{F}$．$n=1,2,\cdots$，则$\bigcup_{n=1}^{\infty}A_n\in\mathscr{F}$.

一般地，称满足上述三个要求的集类为**σ-域**，亦称**σ-代数**．

若$\mathscr{F}$为σ-域，则由(i)及(ii)可得

$$\varnothing\in\mathscr{F}$$

此外，若$A_n\in\mathscr{F}$，$n=1,2,\cdots$，则

$$\bigcap_{n=1}^{\infty} A_n = \overline{\bigcup_{n=1}^{\infty} \bar{A}_n} \in \mathscr{F}$$

$$\bigcup_{n=1}^{k} A_n = A_1 \cup A_2 \cup \cdots \cup A_k \cup \varnothing \cup \cdots \in \mathscr{F}$$

$$\bigcap_{n=1}^{k} A_n = A_1 \cap A_2 \cap \cdots \cap A_k \cap \Omega \cap \Omega \cdots \in \mathscr{F}$$

定义 若 $\mathscr{F}$ 是由样本空间 Ω 的一些子集构成的一个 σ-域，则称它为**事件域**，$\mathscr{F}$ 中的元素称为**事件**，Ω 称为**必然事件**，$\varnothing$ 称为**不可能事件**.

值得指出，按照这种定义，样本点并不一定是事件.

下面我们来举一些例子.

[例 1] $\mathscr{F}=\{\varnothing, \Omega\}$，不难验证 $\mathscr{F}$ 是一个 σ- 域，这时只有必然事件 Ω 与不可能事件 $\varnothing$ 是事件.

[例 2] $\mathscr{F}=\{\varnothing, A, \bar{A}, \Omega\}$. 这时 $\mathscr{F}$ 也是一个 σ-域，Ω, A, $\bar{A}$, $\varnothing$ 是事件.

[例 3] $\Omega=\{\omega_1, \cdots, \omega_n\}$，$\mathscr{F}$ 由 Ω 的一切子集构成，这时 $\mathscr{F}$ 是一个有限的集合，共有 2^n 个元素，不难验证，$\mathscr{F}$ 是一个 σ-域.

[例 4] 对于一般的 Ω，若 $\mathscr{F}$ 由 Ω 的一切子集构成，可以验证 $\mathscr{F}$ 是一个 σ-域.

从上面几个例子中看到，事件域可以很简单，也可以选得十分复杂，这就需要我们根据问题的不同要求来选择适当的事件域.

下面证明若给定 Ω 的一个非空集类 g，必存在唯一的一个 Ω 中的 σ-域 $\mathfrak{M}(g)$，具有如下两个性质：(1)包含 g；(2)若有其他 σ-域包含 g，则必包含 $\mathfrak{M}(g)$. $\mathfrak{M}(g)$ 称为**包含 g 的最小 σ-域**，亦称**由 g 产生的 σ-域**.

先证明必存在包含 g 的 σ-域，显然由 Ω 的一切子集构成的

集类包含了 g，由例 4 知此集类是一个 σ-域，因此至少存在 Ω 中的一个 σ-域 $\mathfrak{M}$，有 $\mathfrak{M}\supset g$.

现在只要取 Ω 中一切包含 g 的 σ-域之交作为 $\mathfrak{M}(g)$，则它是具有上述两个性质的 σ-域. 这点留给读者自行验证.

下面介绍两个非常有用的 σ-域.

[一维波雷尔(Borel)点集] 以后我们将以 R^1 记数直线或实数全体，并称由一切形为 $[a, b)$ 的有界左闭右开区间构成的集类所产生的 σ-域为**一维波雷尔 σ-域**，记之为 $\mathscr{B}_1$，称 $\mathscr{B}_1$ 中的集为**一维波雷尔点集**.

若 x, y 表示任意实数，由于

$$\{x\}=\bigcap_{n=1}^{\infty}\left[x, x+\frac{1}{n}\right)$$

$$(x, y)=[x, y)-\{x\}$$

$$[x, y]=[x, y)+\{y\}$$

$$(x, y]=[x, y)+\{y\}-\{x\}$$

因此 $\mathscr{B}_1$ 中包含一切开区间，闭区间，单个实数，可列个实数，以及由它们经可列次并、交运算而得出的集合. 这是相当大的一个集类，足够把实际问题中感兴趣的点集都包括在内.

[n 维波雷尔点集]以 R^n 记 n 维欧几里得(Euclid)空间，可以类似地定义 n 维波雷尔点集，它们是由一切 n 维矩形产生的 n 维波雷尔 σ-域 $\mathscr{B}_n$ 中的集合，也可以把 R^n 中我们感兴趣的点集都包括在内.

三、概率

在公理化结构中，概率是针对事件定义的，即对应于事件域 $\mathscr{F}$ 中的每一个元素 A 有一个实数 $P(A)$ 与之对应，一般把这种从集合到实数的映照称为集合函数. 因此，概率是定义在事件域 $\mathscr{F}$ 上的一个集合函数. 此外，在公理化结构中只规定概率应满足的性

质，而不具体给出它的计算公式或计算方法.

概率应有什么性质呢？

因为概率通过频率稳定性与随机试验相联系，因此我们自然想到概率应有与频率类似的性质. 关于频率的性质，我们已在§1中总结为非负性，规范性，以及有限可加性.

在古典概型中，概率是通过有利场合数与可能结果总数之比来定义的，它同样具有这三个性质.

在几何概率中，情况也类似，但有一点不同，就是它要求对可列个不相容事件之和有可加性，即可列可加性.

在一般场合，处理可列个事件之和是完全必要的，因此保留这种可列可加性要求看来是合理的.

综上所述，我们如下定义概率.

定义 定义在事件域$\mathscr{F}$上的一个集合函数P称为**概率**，如果它满足如下三个条件：

(i) $P(A)\geqslant 0$，对一切$A\in\mathscr{F}$；

(ii) $P(\Omega)=1$；

(iii) 若$A_i\in\mathscr{F}$，$i=1,2,\cdots$，且两两互不相容，则

$$P\left(\sum_{i=1}^{\infty}A_i\right)=\sum_{i=1}^{\infty}P(A_i) \tag{1}$$

性质(iii)称为**可列可加性**或**完全可加性**.

利用概率的基本性质(i)，(ii)，(iii)可以推出概率的另外一些重要性质.

推论1 不可能事件的概率为0，即$P(\varnothing)=0$.

[证明] 因为

$$\Omega=\Omega+\varnothing+\cdots$$

所以

$$P(\Omega)=P(\Omega)+P(\varnothing)+\cdots$$

因此 $P(\varnothing)=0$.

推论 2 概率具有有限可加性. 即若 $A_iA_j=\varnothing(i\neq j)$, 则

$$P(A_1+A_2+\cdots+A_n)=P(A_1)+P(A_2)+\cdots+P(A_n) \tag{2}$$

[证明] 因为

$$A_1+A_2+\cdots+A_n=A_1+A_2+\cdots+A_n+\varnothing+\varnothing+\cdots$$

由可列可加性及推论 1

$$P(A_1+A_2+\cdots+A_n)=P(A_1)+P(A_2)+\cdots+P(A_n)$$

推论 3 对任何事件 A 有

$$P(\bar{A})=1-P(A) \tag{3}$$

[证明] 因 $A\cup\bar{A}=\Omega, A\bar{A}=\varnothing$, 故

$$1=P(\Omega)=P(A\cup\bar{A})=P(A)+P(\bar{A})$$

推论 4 如果 $A\supset B$, 则

$$P(A-B)=P(A)-P(B) \tag{4}$$

[证明] 因为当 $A\supset B$ 时, $A=B\cup(A-B)$, 又 $B\cap(A-B)=\varnothing$, 故 $P(A)=P(B)+P(A-B)$, 移项即得结论.

系 如果 $A\supset B$, 则 $P(A)\geqslant P(B)$.

推论 5 $P(A\cup B)=P(A)+P(B)-P(AB)$. (5)

[证明] 因 $A\cup B=A\cup(B-AB)$, 而且 $A\cap(B-AB)=\varnothing$, 故 $P(A\cup B)=P(A)+P(B-AB)$, 又 $AB\subset B$, 于是由推论 4 得到

$$P(A\cup B)=P(A)+P(B)-P(AB)$$

系 1 $P(A\cup B)\leqslant P(A)+P(B)$. (6)

系 2 对任意的 n, 成立

$$P(A_1\cup\cdots\cup A_n)\leqslant P(A_1)+\cdots+P(A_n) \tag{7}$$

值得提醒的是概率的这些重要性质的推导实质上只用到概率的非负性, 规范性和有限可加性.

最后，我们对可列可加性作进一步讨论．从推论 2 知道，由可列可加性可以推出有限可加性；但是一般来讲，由有限可加性并不能推出可列可加性．

事实上，若 $A_i\in\mathscr{F}$，$i=1,2,\cdots$，且两两互不相容，则由概率的有限可加性只能推出(2)成立，即

$$P\left(\sum_{i=1}^{n}A_i\right)=\sum_{i=1}^{n}P(A_i) \tag{8}$$

这个等式的左边对任意 n 都不超过 1，因此右边的正项级数收敛．这样应有

$$\lim_{n\to\infty}P\left(\sum_{i=1}^{n}A_i\right)=\lim_{n\to\infty}\sum_{i=1}^{n}P(A_i)=\sum_{i=1}^{\infty}P(A_i) \tag{9}$$

与(1)比较一下可以知道，为了具有可列可加性，还需要下式成立：

$$\lim_{n\to\infty}P\left(\sum_{i=1}^{n}A_i\right)=P\left(\sum_{i=1}^{\infty}A_i\right) \tag{10}$$

或者写成更富有启发性的等式：

$$\lim_{n\to\infty}P\left(\sum_{i=1}^{n}A_i\right)=P\left(\lim_{n\to\infty}\sum_{i=1}^{n}A_i\right) \tag{11}$$

即要允许把极限号通到概率号里面去，这就提出了一个新要求．

现在就来考察这个新要求．若记

$$S_n=\sum_{i=1}^{n}A_i$$

则 $S_n\in\mathscr{F}$，$n=1,2,\cdots$，而且 $S_n\subset S_{n+1}$，即 S_n 是 $\mathscr{F}$ 中一个单调不减的集序列，这时可改写(11)成下式：

$$\lim_{n\to\infty}P(S_n)=P(\lim_{n\to\infty}S_n) \tag{12}$$

一般，对于 $\mathscr{F}$ 上的集合函数，若它对 $\mathscr{F}$ 中任何一个单调不减的

集序列$\{S_n\}$均成立(12)，则我们称它是**下连续的**．因此前面的推导表明，为保证概率的可列可加性成立，除要求它具有有限可加性外，还要求它是下连续的．

下面的定理明确地阐明了这三个概念之间的关系．

定理　若P是$\mathscr{F}$上满足$P(\Omega)=1$的非负集合函数，则它具有可列可加性的充要条件为

(i)　它是有限可加的；

(ii)　它是下连续的．

［证明］　充分性只要重新考察前面的推导过程．实际上，若沿用有关记号，则有限可加性保证了(8)的成立，而下连续性保证了(11)即(10)的成立，通过(9)立刻得到可列可加性．

必要性：其中(i)我们早已建立．为证(ii)，设$\{S_n\}$是$\mathscr{F}$中一个单调不减的集序列，那么

$$\bigcup_{i=1}^{\infty} S_i=\lim_{n\to\infty}S_n$$

若定义$S_0=\varnothing$，则

$$\bigcup_{i=1}^{\infty} S_i=\sum_{i=1}^{\infty}(S_i-S_{i-1})$$

这里的(S_i-S_{i-1})，$i=1, 2, \cdots$，由于S_i的单调性，显然两两互不相容，因此由可列可加性得

$$P\left(\bigcup_{i=1}^{\infty} S_i\right)=\sum_{i=1}^{\infty}P(S_i-S_{i-1})=\lim_{n\to\infty}\sum_{i=1}^{n}P(S_i-S_{i-1})$$

但是

$$\sum_{i=1}^{n}P(S_i-S_{i-1})=P\left(\sum_{i=1}^{n}(S_i-S_{i-1})\right)=P(S_n)$$

因此

$$P(\lim_{n\to\infty} S_n)=\lim_{n\to\infty}P(S_n)$$

这就证得了P的下连续性.

系 1 概率是下连续的.

系 2 概率是上连续的，即若$B_i\in\mathscr{F}$，而且$B_i\supset B_{i+1}$，$i=1, 2, \cdots$，则

$$\lim_{n\to\infty}P(B_n)=P(\lim_{n\to\infty}B_n) \tag{13}$$

［证明］ 记$S_i=\bar{B}_i$，则$\{S_i\}$是单调不减的，由系1可得

$$\lim_{n\to\infty}P(S_n)=P(\lim_{n\to\infty} S_n)=P\left(\bigcup_{i=1}^{\infty}S_i\right)=P\left(\overline{\bigcap_{i=1}^{\infty}B_i}\right)$$

因此由(3)，可知

$$1-\lim_{n\to\infty}P(B_n)=1-P\left(\bigcap_{i=1}^{\infty}B_i\right)$$

注意到$\lim_{n\to\infty} B_n=\bigcap_{i=1}^{\infty}B_i$，即得(13).

四、概率空间

在柯尔莫哥洛夫的概率论公理化结构中，称三元总体$(\Omega, \mathscr{F}, P)$为概率空间，其中Ω是样本空间，$\mathscr{F}$是事件域，P是概率，它们都认为是预先给定的，并以此作为出发点讨论种种问题．至于实际问题中，如何选定Ω，怎样构造$\mathscr{F}$，怎样给定P，则要视具体情况而定.

下面讨论几个具体例子.

［例5］（有限概率空间）这时Ω中只有n个点，这类概率空间在§2中已有讨论，古典概型即为其特例．在这种场合，一般可取$\mathscr{F}$为Ω的所有子集全体，这仍是一个有限的集合，元素总数为2^n，它满足事件域的三个要求，而且样本点（看作一个单点集）是事件.

至于概率，只要对样本点 ω_i 给定满足

$$P(\omega_i)\geqslant 0,\quad i=1,2,\cdots,n \tag{14}$$

$$P(\omega_1)+\cdots+P(\omega_n)=1 \tag{15}$$

的一组数 $P(\omega_1),P(\omega_2),\cdots,P(\omega_n)$，那么，若 A 是 $\mathscr{F}$ 中元素，包含样本点 $\omega_{i_1},\cdots,\omega_{i_k}$，则由概率的可加性，自然应令

$$P(A)=P(\omega_{i_1})+\cdots+P(\omega_{i_k})$$

这就给定了事件 A 的概率，从而构成了概率空间 $(\Omega,\mathscr{F},P)$.

从这个例子中看到下面两点.

(1) 选定了 $(\Omega,\mathscr{F})$ 之后，对于事件概率的给定还有相当大的灵活性，这表现在 $P(\omega_i)$ 的选取上. 因为只有这样，才能用概率空间来描述不同的随机现象. 例如在投一次硬币的试验中，Ω 总是由出正面 (ω_1) 及出反面 (ω_2) 两个样本点构成. 对于均匀的硬币，可以假定它出正面及反面的概率均为 $\frac{1}{2}$；但对于很不均匀的硬币，例如出正面可能性大得多的硬币，则必须给定另外的概率，而这只须适当给定 $P(\omega_1)$ 就可以了.

(2) 一旦 $P(\omega_i)$ 给定后，事件 A 的概率并不能任意给定，即在事件域中，各事件的概率有一定关系，给定概率时必须满足这些关系.

[例 6] (离散概率空间) Ω 由可列个点构成：$\Omega=\{\omega_1,\ \omega_2,\ \cdots\}$，这时 $\mathscr{F}$ 还是可以选为 Ω 的子集全体，它满足事件域的三个条件. 这时样本点也是事件. 为给定概率，可选可列个非负的数 P_i，$i=1,2,\cdots$，满足

$$\sum_{i=1}^{\infty}P_i=1$$

作为样本点 ω_i 的概率，而一般事件 $A\in\mathscr{F}$ 的概率，则必须取为它所含的样本点的概率之和.

[例 7] 若 $\Omega=R^1$，即样本空间由全体实数构成，这时$\mathscr{F}$不能取为Ω的一切子集，因为这个集类太大，无法在其上定义概率. 这时通常取$\mathscr{F}$为直线上波雷尔点集全体 $\mathscr{B}_1$，这是相当大的一个集类，可以把实际问题中所感兴趣的点集都包括在内. 另一方面在波雷尔点集上定义概率相当方便，这只要对左闭右开区间给定概率即可. 这些我们将在第三章中作深入讨论.

顺便指出，若Ω不是 R^1，而是它的一部分，也可类似处理. 譬如Ω为一个区间，这时$\mathscr{F}$可取为该区间上的波雷尔点集全体：它们通过直线上波雷尔点集与该区间之交而得到.

[例 8] 若 $\Omega=R^n$(或 R^n 的一部分)，这时可类似于一维场合取 n 维欧几里得空间中的波雷尔点集全体 $\mathscr{B}_n$ 作为事件域$\mathscr{F}$，在第三章 § 2 中将对这种场合进行深入讨论.

第一章小结

本章中介绍了一类新的现象——随机现象，这是一种普遍存在的现象. 在大量随机现象中存在着统计规律性，概率论便是研究随机现象的数量规律的一门数学学科.

“事件”与“概率”是概率论中最基本的两个概念，我们在公理化结构下严格地定义了这两个概念.

为了使读者清楚地理解事件与概率的直观意义，我们采用由具体到抽象，由简单到复杂，由特殊到一般的方式分别介绍了频率、古典概型、几何概率，并从中归纳出事件与概率的本质特征，为公理化定义作准备，这种讲法基本上与概率概念的历史发展平行.

事件的运算及概率的性质是本章的基本内容，也是学习以后各章的必要基础，务必牢固掌握.

我们较完整地研究了古典概型，并介绍了它在产品抽样检查中的应用. 对于古典概型的讨论有助于概率论基本概念的直观理

解，而且在以后讨论更一般的情况时，也常以它为特例加以考虑.

古典概型中概率的计算有较高的技巧性，同学们应该掌握住一些最基本的计算方法.

几何概率是很有启发性的一类问题，不过它的严格表述只有用到第三章的一些概念才能做到.

搞清频率与概率的关系是十分重要的一个课题，今后我们将一再回到这个问题上来.

书中我们着重把现代概率论公理化结构作为一个历史发展过程来描述，有兴趣的读者不妨试着进行相反而相成的另一工作：用公理化结构来概括古典概型、几何概率等特殊模型.

波雷尔点集是一个重要的概念，在第三章中，它将起关键作用.

习　题

1．在某城市中，共发行三种报纸 A, B, C. 在这城市的居民中，订购 A 的占45%，订购 B 的占35%，订购 C 的占30%，同时订购 A 及 B 的占10%，同时订购 A 及 C 的占8%，同时订购 B 及 C 的占5%，同时订购 A, B, C 的占3%，试求下列百分率：(1)只订购 A 的；(2)只订购 A 及 B 的；(3)只订购一种报纸的；(4)正好订购两种报纸的；(5)至少订购一种报纸的；(6)不订购任何报纸的.

2．若 A, B, C 是随机事件，说明下列关系式的概率意义：(1)$ABC=A$；(2) $A\cup B\cup C=A$；(3)$AB\subset C$；(4)$A\subset\overline{BC}$.

3．试把 $A_1\cup A_2\cup\cdots\cup A_n$ 表示成 n 个两两互不相容事件的和.

4．在某班学生中任选一个同学，以事件 A 表示选到的是男同学，事件 B 表示选到的人不喜欢唱歌，事件 C 表示选到的人是运动员．(1)表述 $AB\bar{C}$ 及 $A\bar{B}C$；(2)什么条件下成立 $ABC=A$；(3) 何时成立 $\bar{C}\subset B$；(4) 何时同时成立 $A=B$ 及 $\bar{A}=C$.

5．用摸球模型造一例，指出样本空间及各种事件运算.

6．若 A, B, C, D 是四个事件，试用这四个事件表示下列各事件：(1)这

四个事件至少发生一个；(2)这四个事件恰好发生两个；(3)A, B 都发生而 C, D 都不发生；(4) 这四个事件都不发生；(5) 这四个事件中至多发生一个。

*7. 从 0, 1, 2, ⋯, 9 中随机地取出 5 个数(可重复), 以 E_i 记某些数正好出现 i 次这一事件(例如 52353, 既属于 E_1 也属于 E_2 及 E_0), 试用文图表示 $E_0, E_1, \cdots, E_5$ 的关系.

8. 证明下列等式:

(1) $\binom{n}{1}+2\binom{n}{2}+3\binom{n}{3}+\cdots+n\binom{n}{n}=n2^{n-1}$

(2) $\binom{n}{1}-2\binom{n}{2}+3\binom{n}{3}+\cdots+(-1)^{n-1}n\binom{n}{n}=0$

(3) $\sum_{k=0}^{a-r}\binom{a}{k+r}\binom{b}{k}=\binom{a+b}{a-r}$

9. 袋中有白球 5 只, 黑球 6 只, 陆续取出三球, 求顺序为黑白黑的概率.

10. 一部五本头的文集, 按任意次序放到书架上去, 试求下列概率: (1) 第一卷出现在旁边; (2) 第一卷及第五卷出现在旁边; (3) 第一卷或第五卷出现在旁边; (4)第一卷及第五卷都不出现在旁边; (5)第三卷正好在正中.

11. 把 1, 2, 3, 4, 5 诸数各写在一小纸片上, 任取其三而排成自左向右的次序, 求所得数是偶数的概率.

12. 在一个装有 n 只白球, n 只黑球, n 只红球的袋中, 任取 m 只球, 求其中白、黑、红球分别有 $m_1, m_2, m_3(m_1+m_2+m_3=m)$ 只的概率.

13. 甲袋中有 3 只白球, 7 只红球, 15 只黑球, 乙袋中有 10 只白球, 6 只红球, 9 只黑球, 现从两袋中各取一球, 求两球颜色相同的概率.

14. 由盛有号码 1, 2, ⋯, N 的球的箱子中有放回地摸了 n 次球, 依次记下其号码, 试求这些号码按严格上升次序排列的概率.

*15. 在上题中这些号码按上升(不一定严格)次序排列的概率.

16. 任意从数列 1, 2, ⋯, N 中不放回地取出 n 个数并按大小排列成: $x_1<x_2<\cdots<x_m<\cdots<x_n$, 试求 $x_m=M$ 的概率. 这里 $1\leqslant M\leqslant N$.

*17. 上题中, 若采用有放回取数, 这时 $x_1\leqslant x_2\leqslant\cdots\leqslant x_n$, 试求 $x_m=M$ 的概率.

18. 从 6 双不同的手套中任取 4 只, 问其中恰有一双配对的概率是多少?

19. 从 n 双不同的鞋子中任取 $2r(2r<n)$ 只，求下列事件发生的概率：(1)没有成对的鞋子；(2)只有一对鞋子；(3) 恰有两对鞋子；(4) 有 r 对鞋子。

20. 袋中有 n 只球，记有号码 1, 2, …, n，求下列事件的概率：(1)任意取出 2 球，号码为 1, 2；(2)任意取出 3 球，没有号码 1；(3)任意取出 5 球，号码 1, 2, 3 中至少出现一个。

21. 袋中装有 1, 2, …, N 号的球各一只，采用(1)有放回；(2)不放回方式摸球，试求在第 k 次摸球时首次摸到 1 号球的概率。

22. 甲有 $n+1$ 个硬币，乙有 n 个硬币，双方投掷之后进行比较，求甲掷出的正面比乙掷出的正面多的概率。

23. (De Méré 问题)一颗骰子投 4 次至少得到一个六点与两颗骰子投 24 次至少得到一个双六这两件事哪一件有更多的机会遇到？

24. 从 52 张扑克牌中任意取出 13 张来，问有 5 张黑桃，3 张红心，3 张方块，2 张草花的概率。

25. 桥牌游戏中(四人各从 52 张纸牌中分得 13 张)，求 4 张 A 集中在一个人手中的概率。

26. 在扑克牌游戏中(从 52 张牌中任取 5 张)，求下列事件的概率：(1)以 A 打头的同花顺次五张牌；(2) 其他同花顺次五张牌；(3) 有四张牌同点数；(4) 三张同点数且另两张也同点数；(5) 五张同花；(6) 异花顺次五张牌；(7)三张同点数，另外两张不同点数；(8)五张中有两对；(9)五张中有一对；(10)其他情况。

27. 某码头只能容纳一只船，现预知某日将独立来到两只船，且在 24 小时内各时刻来到的可能性都相等，如果它们需要停泊的时间分别为 3 小时及 4 小时，试求有一船要在江中等待的概率。

28. 两人约定于 7 点到 8 点在某地会面，试求一人要等另一人半小时以上的概率。

29. 在一线段 AB 中随机地取两个点 X_1 与 X_2，求 AX_1, X_1X_2, X_2B 可以构成一个三角形的概率。(三线段能构成三角形的充要条件是任意二边之和大于第三边)

30. 在线段[0, 1]上任意投三个点，问由 0 至三点的三线段能构成三角形与不能构成三角形这两个事件中哪一个事件的概率大。

31. 在一张打上方格的纸上投一枚直径为 1 的硬币，方格要多小才能使

硬币与线不相交的概率小于1%.

32. 从(0, 1)中随机地取两个数，求下列概率：(1)两数之和小于1.2；(2)两数之积小于$\frac{1}{4}$；(3)以上两个要求同时满足.

33. 设 $A_1, A_2, \cdots, A_n$ 是随机事件，试用归纳法证明下列公式

$$P(A_1 \cup A_2 \cup \cdots \cup A_n)=\sum_{i=1}^{n}P(A_i)-\sum_{1\leqslant i<j\leqslant n}P(A_iA_j)$$
$$+\sum_{1\leqslant i<j<k\leqslant n}P(A_iA_jA_k)-\cdots$$
$$+(-1)^{n-1}P(A_1A_2\cdots A_n)$$

34. 某班有N个士兵，每人各有一支枪，这些枪外形完全一样，在一次夜间紧急集合中，若每人随机地取走一支枪，问至少有一个人拿到自己的枪的概率.

*35. 在上题中求恰好有 $k(0\leqslant k\leqslant N)$个人拿到自己的枪的概率.

36. 考试时共有N张考签，n个学生参加考试($n\geqslant N$)，被抽过的考签立刻放回，求在考试结束之后，至少有一张考签没有被抽到的概率.

37. 甲，乙，丙三人按下面规则进行比赛，第一局由甲，乙参加而丙轮空，由第一局的优胜者与丙进行第二局比赛，而失败者则轮空，比赛用这种方式一直进行到其中一个人连胜两局为止，连胜两局者成为整场比赛的优胜者，若甲，乙，丙胜每局的概率各为$\frac{1}{2}$，问甲，乙，丙成为整场比赛优胜者的概率各是多少？

38. 父，母，子三人举行比赛，每局总有一人胜一人负(没有和局)，每局的优胜者就与未参加此局的人再进行比赛，如果某人首先打胜了两局，则他就是整个比赛的优胜者，由父决定第一局由哪两人参加，其中儿子实力最强，所以父为了使自己得胜的概率达到最大，就决定第一局由他与妻子先比赛，试证父的决策为最优策略(任何一对选手中一人胜对方的概率在整个比赛中是不变的).

39. 给定 $p=P(A)$，$q=P(B)$，$r=P(A\cup B)$，求 $P(A\bar{B})$及 $P(\bar{A}\bar{B})$.

40. 已知：$P(AB)=P(A)P(B)$，$C\supset AB$，$\bar{C}\supset\bar{A}\bar{B}$，证明：$P(AC)\geqslant P(A)P(C)$.

41. (1) 已知 A_1 与 A_2 同时发生则 A 发生，试证 $P(A)\geqslant P(A_1)+P(A_2)-1$.

(2) 若 $A_1A_2A_3\subset A$，试证 $P(A)\geqslant P(A_1)+P(A_2)+P(A_3)-2$.

*42. 用概率论想法求N阶行列式的展开式中包含主对角线元素的项数.

43. 利用概率论的想法证明下列恒等式:

$$1+\frac{A-a}{A-1}+\frac{(A-a)(A-a-1)}{(A-1)(A-2)}+\cdots+\frac{(A-a)\cdots2\cdot1}{(A-1)\cdots(a+1)a}=\frac{A}{a}$$

其中 A,a 都是正整数，且 $A>a$.

44. 用某种药物对患有胃溃疡的512个病人进行治疗，结果368人有明显疗效，现有某胃溃疡病人欲服此药，你能对其效果作何估计？

45. 求包含事件 A,B 的最小 σ-域.

46. 证明：Ω 的一切子集组成的集类是一个 σ-域.

47. 证明：σ-域之交仍为 σ-域.

48. 证明：包含一切形为$(-\infty,x)$的区间的最小 σ-域是一维波雷尔 σ-域.

49. 设Q是定义在 σ-域上的非负广义实值函数（即可以取有限或无限值的函数），如果它具有可列可加性，并且 $Q(\varnothing)=0$，则称Q为测度．试说明测度概念是算术中计数概念及几何中长度、面积、体积等概念的推广.

50. 用测度概念解释古典概型、几何概率及概率论公理化结构中关于概率的定义.

第二章　条件概率与统计独立性

§1. 条件概率，全概率公式，贝叶斯公式

一、条件概率

对概率的讨论总是在一组固定的条件限制下进行的．以前的讨论总是假定除此之外再无别的信息可供使用．可是，有时我们却会碰到这样的情况，即已知某一事件B已经发生，要求另一事件A发生的概率．例如考虑有两个孩子的家庭，假定男女出生率一样，则两个孩子（依大小排列）的性别为（男，男），（男，女），（女，男），（女，女）的可能性是一样的．若以A记随机选取的这样一个家庭中有一男一女这一事件，则显然$P(A)=\frac{1}{2}$，但是如果我们预先知道这家庭至少有一个女孩，那么，上述事件的概率便应是$\frac{2}{3}$．

两种情况下算出的概率不同．这也很容易理解，因为在第二种情况下，我们多知道了一个条件：事件B（这一家庭至少有一女孩）发生，因此我们算得的概率事实上是"在已知事件B发生的条件下，事件A发生的概率"，这个概率我们将记之为$P(A|B)$．

这种带有条件的概率很重要，下面我们就来研究它．在给出严格定义之前，先考察一些特殊的场合．

就从上述例子出发．这是一个古典概型问题，样本点总数$n=4$，有利于事件A的场合数$m_A=2$，因此$P(A)=\frac{1}{2}$；但是假如已知事件B发生，即至少有一女孩，那么可能发生的样本点是（男，女），（女，男），（女，女），总数为$m_B=3$，而有利场合（至少有一女孩

而且有一男一女)数 $m_{AB}=2$,因此

$$P(A|B)=\frac{2}{3}=\frac{m_{AB}}{m_B}=\frac{\frac{m_{AB}}{n}}{\frac{m_B}{n}}=\frac{P(AB)}{P(B)} \tag{1}$$

这式子很重要,虽然我们以特例形式引入,但读者不难证明,它对一般古典概型问题也成立.

在几何概率中,若以 $m(A)$, $m(B)$, $m(AB)$, $m(\Omega)$分别记事件 A, B, AB, Ω 所对应点集的测度,且 $m(B)>0$,则

$$P(A|B)=\frac{m(AB)}{m(B)}=\frac{\frac{m(AB)}{m(\Omega)}}{\frac{m(B)}{m(\Omega)}}=\frac{P(AB)}{P(B)}$$

结果与古典概型相同.

对频率也有类似结果,请读者自行验证.

在一般场合,我们将把这个算式作为条件概率的定义.

定义 设$(\Omega, \mathscr{F}, P)$是一个概率空间,$B\in\mathscr{F}$,而且$P(B)>0$,则对任意 $A\in\mathscr{F}$,记

$$P(A|B)=\frac{P(AB)}{P(B)} \tag{2}$$

并称 $P(A|B)$ 为**在事件 B 发生的条件下事件 A 发生的条件概率.**

由(2)立刻得到

$$P(AB)=P(B)P(A|B) \tag{3}$$

这个等式有时被称为概率的**乘法定理**.

若还有 $P(A)>0$,则也可定义 $P(B|A)$,这时有

$$P(AB)=P(A)P(B|A)=P(B)P(A|B) \tag{4}$$

[例 1] 甲,乙两市都位于长江下游,根据一百多年来的气象记录,知道一年中雨天的比例甲市占 20%,乙市占 18%,两地同时下雨占 12%。若以事件 A 记甲市出现雨天,事件 B 记乙市出现雨

天，则$P(A)=0.20$, $P(B)=0.18$, $P(AB)=0.12$，因此

$$P(A|B)=\frac{P(AB)}{P(B)}=\frac{0.12}{0.18}=0.67$$

$$P(B|A)=\frac{P(AB)}{P(A)}=\frac{0.12}{0.20}=0.60$$

此外

$$P(A\cup B)=P(A)+P(B)-P(AB)=0.20+0.18-0.12$$
$$=0.26$$

看来，在下雨这件事情上，应认为甲乙两市是有联系的。

下面讨论条件概率的性质。

首先，不难验证条件概率$P(A|B)$具有概率的三个基本性质：非负性，规范性，完全可加性。

(i) $P(A|B)\geqslant 0$;

(ii) $P(\Omega|B)=1$;

(iii) $P\left(\sum_{i=1}^{\infty}A_i|B\right)=\sum_{i=1}^{\infty}P(A_i|B)$.

因此，类似于概率，对条件概率也可由三个基本性质导出其它一些性质，例如

$$P(\varnothing|B)=0$$
$$P(A|B)=1-P(\bar{A}|B)$$
$$P(A_1\cup A_2|B)=P(A_1|B)+P(A_2|B)-P(A_1A_2|B)$$

特别当$B=\Omega$时，条件概率化为无条件概率，因此把一般的概率看作条件概率也未尝不可。

可以把乘法定理推广到任意n个事件之交的场合：

$$P(A_1A_2\cdots A_n)=P(A_1)P(A_2|A_1)P(A_3|A_1A_2)\cdots P(A_n|A_1A_2\cdots A_{n-1}) \tag{5}$$

这里当然要求$P(A_1A_2\cdots A_{n-1})>0$.

［例 2］ 卜里耶(Pólya)模型罐中有b只黑球及r只红球，随机取出一只，把原球放回，并加进与抽出球同色的球c只，再摸第二次，这样下去共摸了n次，问前面的n_1次出现黑球，后面的$n_2=n-n_1$次出现红球的概率是多少？

［解］ 以A_1表示第一次摸出黑球这一事件，……，A_{n_1}表示第n_1次摸出黑球，A_{n_1+1}表示第n_1+1次摸出红球，……，A_n表示第n次摸出红球．则

$$P(A_1)=\frac{b}{b+r},\qquad P(A_2|A_1)=\frac{b+c}{b+r+c}$$

$$P(A_3|A_1A_2)=\frac{b+2c}{b+r+2c},\cdots$$

$$P(A_{n_1}|A_1\cdots A_{n_1-1})=\frac{b+(n_1-1)c}{b+r+(n_1-1)c}$$

$$P(A_{n_1+1}|A_1\cdots A_{n_1})=\frac{r}{b+r+n_1c}$$

$$P(A_{n_1+2}|A_1\cdots A_{n_1+1})=\frac{r+c}{b+r+(n_1+1)c},\cdots$$

$$P(A_n|A_1\cdots A_{n-1})=\frac{r+(n_2-1)c}{b+r+(n-1)c}$$

因此

$$P(A_1A_2\cdots A_n)=\frac{b}{b+r}\cdot\frac{b+c}{b+r+c}\cdot\frac{b+2c}{b+r+2c}\cdots\frac{b+(n_1-1)c}{b+r+(n_1-1)c}\cdot\frac{r}{b+r+n_1c}\cdot\frac{r+c}{b+r+(n_1+1)c}\cdots\frac{r+(n_2-1)c}{b+r+(n-1)c}$$

注意这个答案只与黑球及红球出现次数有关，而与出现的顺序无关．

这个模型曾被卜里耶用来作为描述传染病的数学模型．这是很一般的摸球模型，特别取$c=0$，则是有放回摸球，取$c=-1$，则

是不放回摸球.

二、全概率公式

概率论的重要研究课题之一是希望从已知的简单事件的概率推算出未知的复杂的事件的概率. 为达到这个目的, 经常把一个复杂事件分解为若干个不相容的简单事件之和, 再通过分别计算这些简单事件的概率, 最后利用概率的可加性得到最终结果. 这里, 全概率公式起着很重要的作用.

设事件 $A_1, A_2, \cdots, A_n, \cdots$ 是样本空间 Ω 的一个分割, 即 A_i 两两互不相容, $P(A_i)>0$, $i=1,2,\cdots$, 而且

$$\sum_{i=1}^{\infty} A_i = \Omega$$

这样一来

$$B = \sum_{i=1}^{\infty} A_i B$$

这里的 A_iB 也两两互不相容(参看图 1).

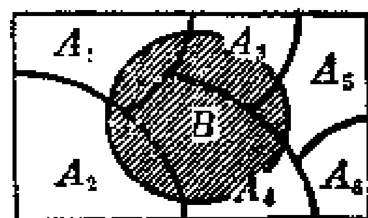


图 1 样本空间的分割

由概率的完全可加性

$$P(B) = \sum_{i=1}^{\infty} P(A_i B)$$

再利用乘法定理即得

$$P(B) = \sum_{i=1}^{\infty} P(A_i) P(B | A_i) \tag{6}$$

这公式称为**全概率公式**, 它是概率论的一个基本公式, 有着多方面的应用.

[例 3] 播种用的一等小麦种子中混合 2% 的二等种子, 1.5% 的三等种子, 1% 的四等种子. 用一等、二等、三等、四等种子长出的穗含 50 颗以上麦粒的概率分别是 0.5, 0.15, 0.1, 0.05, 求这

批种子所结的穗含有50颗以上麦粒的概率.

[解] 设从这批种子中任选一颗是一等、二等、三等、四等种子的事件分别记为A_1, A_2, A_3, A_4, 则它们构成样本空间的一个分割. 用B表示在这批种子中任选一颗, 且这颗种子所结的穗含有50颗以上麦粒这一事件, 则由全概率公式得

$$P(B)=\sum_{i=1}^{4}P(A_i)P(B|A_i)$$
$$=95.5\%\times0.5+2\%\times0.15+1.5\%\times0.1+1\%\times0.05$$
$$=0.4825$$

三、贝叶斯(Bayes)公式

若事件B能且只能与两两互不相容的事件$A_1, A_2, \cdots, A_n, \cdots$之一同时发生,即

$$B=\sum_{i=1}^{\infty}BA_i$$

由于

$$P(A_iB)=P(B)P(A_i|B)=P(A_i)P(B|A_i)$$

故

$$P(A_i|B)=\frac{P(A_i)P(B|A_i)}{P(B)}$$

再利用全概率公式即得

$$P(A_i|B)=\frac{P(A_i)P(B|A_i)}{\sum_{i=1}^{\infty}P(A_i)P(B|A_i)} \tag{7}$$

这个公式称为贝叶斯公式.

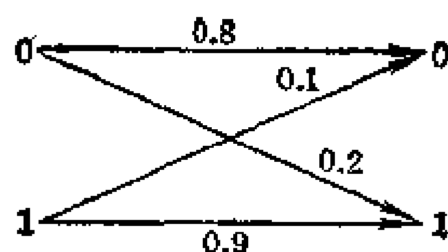
贝叶斯公式在概率论和数理统计中有着多方面的应用. 假定$A_1, A_2, \cdots$是导致试验结果的"原因", $P(A_i)$称为**先验概率**, 它反映了各种"原因"发生的可能性大小, 一般是以往经验的总结, 在这次

试验前已经知道．现在若试验产生了事件B，这个信息将有助于探讨事件发生的“原因”．条件概率$P(A_i|B)$称为**后验概率**，它反映了试验之后对各种“原因”发生的可能性大小的新知识．例如在医疗诊断中，有人为了诊断病人到底是患了毛病$A_1,A_2,\cdots,A_n$中的哪一种，对病人进行观察与检查，确定了某个指标B(譬如是体温、脉搏、血液中转氨酶含量等等)，他想用这类指标来帮助诊断．这时就可以用贝叶斯公式来计算有关概率．首先必须确定先验概率$P(A_i)$，这实际上是确定人患各种毛病的可能性大小，以往的资料可以给出一些初步数据；其次是要确定$P(B|A_i)$，这里当然主要依靠医学知识．有了它们，利用贝叶斯公式就可算出$P(A_i|B)$，显然，对应于较大$P(A_i|B)$的“病因”A_i，应多加考虑．在实际工作中，检查的指标B一般有多个，综合所有的后验概率，当然会对诊断有很大帮助．在实现计算机自动诊断或辅助诊断中，这方法是有实用价值的．

下面介绍应用贝叶斯公式的几个例子．

[例4]　在数字通信中，由于存在着随机干扰，因此接收到的信号与发出的信号可能不同，为了确定发出的信号，通常要计算各种概率，下面只讨论一种比较简单的模型——二进信道．

若发报机以0.7和0.3的概率发出信号0和1(譬如分别用低电平与高电平表示)，由于随机干扰的影响，当发出信号0时，接收机不一定收到0，而是以概率0.8和0.2收到信号0和1；同样地，当发报机发出信号1时，接收机以概率0.9和0.1收到信号1和0．其关系如右图所示．



假如接收机收到信号0，则这时有两种可能情况：第一种情况是发报机确实发出信号0，而信号传输中没

有出错；第二种情况是发报机发出信号 1，但是传输中出现错误，因此接收到信号 0. 所以在这类问题中，计算“当接收机收到信号 0 时，发报机是发出信号 0 的概率”便很有必要.

把发报机发出信号 0 记为事件 A_0，发出信号 1 记为事件 A_1，接收机接到信号 0 记为 B，我们要求的是 $P(A_0|B)$.

由于 $P(A_0)=0.7, P(A_1)=0.3, P(B|A_0)=0.8, P(B|A_1)=0.1$，用贝叶斯公式

$$P(A_0|B)=\frac{P(A_0)P(B|A_0)}{P(A_0)P(B|A_0)+P(A_1)P(B|A_1)}$$

$$=\frac{0.7\times0.8}{0.7\times0.8+0.3\times0.1}=\frac{0.56}{0.59}=0.949$$

［例 5］ 假定用血清甲胎蛋白法诊断肝癌，$P(A|C)=0.95$，$P(\bar{A}|\bar{C})=0.90$，这里 C 表示被检验者患有肝癌这一事件，A 表示判断被检验者患有肝癌这一事件. 又设在人群中 $P(C)=0.0004$. 现在若有一人被此检验法诊断为患有肝癌，求此人真正患有肝癌的概率 $P(C|A)$.

［解］ 由贝叶斯公式

$$P(C|A)=\frac{P(C)P(A|C)}{P(C)P(A|C)+P(\bar{C})P(A|\bar{C})}$$

$$=\frac{0.0004\times0.95}{0.0004\times0.95+0.9996\times0.1}=0.0038$$

因此，虽然检验法相当可靠，但是被诊断为肝癌的人确实患有肝癌的可能性并不大.

［例 6］ （贝叶斯决策）为了判断一根木材是桦木还是桉木，通常采用先抽取它的某一个特征（例如平均亮度 X），然后再根据这个特征作出判决，这时贝叶斯决策是常用方法之一.

以 A_1，A_2 分别记被检验的木材为桦木或桉木这一事件，它们的先验概率 $P(A_1)$ 及 $P(A_2)$ 应预先给定，此外要通过试验确定

$P(X|A_1)$及$P(X|A_2)$，由贝叶斯公式得

$$P(A_i|X)=\frac{P(A_i)P(X|A_i)}{\sum_{i=1}^{2}P(A_i)P(X|A_i)}$$

其中，$i=1,2$，若$P(A_1|X)>P(A_2|X)$，则作出决策：具有特征X的木材是桦木.

这个方法在模式识别这一新兴学科中有重大应用，当然，这里是大为简化了的模型.

§2. 事件独立性

一、两个事件的独立性

本节中我们引进一个新的概念——统计独立性．先从两个事件的独立性开始，然后讨论更为一般的场合.

还是从考虑古典概型的一个例子作为出发点.

［例1］ 一袋中装有a只黑球和b只白球，采用有放回摸球，求：

(1) 在已知第一次摸得黑球的条件下，第二次摸出黑球的概率；

(2) 第二次摸出黑球的概率.

［解］ 以事件A表示第一次摸得黑球，事件B表示第二次摸得黑球，则

$$P(A)=\frac{a}{a+b},P(AB)=\frac{a^2}{(a+b)^2},P(\bar{A}B)=\frac{ba}{(a+b)^2}$$

所以

$$P(B|A)=\frac{P(AB)}{P(A)}=\frac{a}{a+b}$$

而

$$P(B)=P(AB)+P(\bar{A}B)=\frac{a^2}{(a+b)^2}+\frac{ba}{(a+b)^2}=\frac{a}{a+b}$$

注意这里的 $P(B|A)=P(B)$，即事件 A 发生与否，对事件 B 发生的概率没有影响．从直观上讲，这很自然．因为我们这里采用的是有放回摸球，因此第二次摸球时袋中球的组成与第一次摸球时完全相同，当然第一次摸球的结果实际上不影响第二次摸球．在这种场合可以说，事件 A 与事件 B 的出现有某种"独立性"．

对此，我们引进

定义 对事件 A 及 B，若

$$P(AB)=P(A)P(B) \tag{1}$$

则称它们是**统计独立的**，简称**独立的**．

注意，按照这个定义，必然事件 Ω 及不可能事件 $\varnothing$ 与任何事件独立．此外，从(1)中看出，A 与 B 的位置对称，因此亦称 **A 与 B 相互独立**．

推论 1 若事件 A, B 独立，且 $P(B)>0$，则

$$P(A|B)=P(A) \tag{2}$$

［证明］ 由条件概率定义及(1)得

$$P(A|B)=\frac{P(AB)}{P(B)}=\frac{P(A)P(B)}{P(B)}=P(A)$$

因此，若事件 A, B 相互独立，则 A 关于 B 的条件概率等于无条件概率 $P(A)$，这表示 B 的发生对于事件 A 是否发生没有提供任何消息，独立性就是把这种关系从数学上加以严格定义．

推论 2 若事件 A 与 B 独立，则下列各对事件也相互独立：

$$\{\bar{A}, B\},\quad \{A, \bar{B}\},\quad \{\bar{A}, \bar{B}\}$$

［证明］ 由于

$$\begin{aligned} P(\bar{A}B) &= P(B-AB)=P(B)-P(AB) \\ &= P(B)-P(A)P(B)=P(B)[1-P(A)] \\ &= P(\bar{A})P(B) \end{aligned}$$

所以 $\bar{A}$ 与 B 相互独立，由它立刻推出 $\bar{A}$ 与 $\bar{B}$ 相互独立，由 $\bar{\bar{A}}=A$

又推出 $A, \bar{B}$ 相互独立.

不放回摸球模型提供了不独立的一个简单的例子.

[例 2] 在前例中,若采用不放回摸球,试求同样那两个事件的概率.

[解] 这时

$$P(A)=\frac{a}{a+b}, \qquad P(AB)=\frac{a(a-1)}{(a+b)(a+b-1)}$$

$$P(\bar{A}B)=\frac{ba}{(a+b)(a+b-1)}$$

所以

$$P(B|A)=\frac{P(AB)}{P(A)}=\frac{a-1}{a+b-1}$$

而

$$P(B)=P(AB)+P(\bar{A}B)=\frac{a}{a+b}$$

这里 $P(B|A)\neq P(B)$, 即事件 B 与事件 A 不是相互独立的. 因为第一次摸得黑球,事实上已使袋中球的组成成份改变了,当然要影响第二次摸得黑球的概率.

二、多个事件的独立性

我们先定义三个事件 A, B, C 的独立性.

定义 对于三个事件 A, B, C, 若下列四个等式同时成立, 则称它们相互独立.

$$\left.\begin{aligned} P(AB)&=P(A)P(B)\\ P(BC)&=P(B)P(C)\\ P(AC)&=P(A)P(C)\end{aligned}\right\} \tag{3}$$

$$P(ABC)=P(A)P(B)P(C) \tag{4}$$

按两个事件独立性的定义, 我们知道若(3)成立, 则 A 与 B, B 与 C, C 与 A 都相互独立, 也即 A, B, C 两两独立.

读者自然会提出这样一个问题：三个事件 A, B, C 两两独立，能否保证它们相互独立呢？即能否由(3)推出(4)？回答是否定的，这从下面简单的例子就可看出．

［例 3］ 一个均匀的正四面体，其第一面染成红色，第二面染成白色，第三面染成黑色，而第四面同时染上红，白，黑三种颜色．现在我们以 A, B, C 分别记投一次四面体出现红，白，黑颜色的事件，则由于在四面体中有两面有红色，因此

$$P(A)=\frac{1}{2}$$

同理 $P(B)=P(C)=\frac{1}{2}$，容易算出

$$P(AB)=P(BC)=P(AC)=\frac{1}{4}$$

所以(3)成立，也即 A, B, C 两两独立，但是

$$P(ABC)=\frac{1}{4}\neq\frac{1}{8}=P(A)P(B)P(C)$$

因此(4)不成立，从而 A, B, C 不相互独立．

下面再提供一个例子说明由(4)不能推出(3)，进一步说明要 A, B, C 相互独立必须同时要求(3)及(4)成立．

［例 4］ 若有一个均匀正八面体，其第 1, 2, 3, 4 面染红色，第 1, 2, 3, 5 面染白色，第 1, 6, 7, 8 面染上黑色，现在以 A, B, C 分别表示投一次正八面体出现红，白，黑的事件，则

$$P(A)=P(B)=P(C)=\frac{4}{8}=\frac{1}{2}$$

$$P(ABC)=\frac{1}{8}=P(A)P(B)P(C)$$

但是

$$P(AB)=\frac{3}{8}\neq\frac{1}{4}=P(A)P(B)$$

现在我们可以定义 n 个事件的独立性.

定义 对 n 个事件 $A_1, A_2, \cdots, A_n$，若对于所有可能的组合 $1\leqslant i<j<k<\cdots\leqslant n$ 成立着

$$\left.\begin{array}{l} P(A_iA_j)=P(A_i)P(A_j) \\ P(A_iA_jA_k)=P(A_i)P(A_j)P(A_k) \\ \cdots \\ P(A_1A_2\cdots A_n)=P(A_1)P(A_2)\cdots P(A_n) \end{array}\right\} \tag{5}$$

则称 $A_1, A_2, \cdots, A_n$ 相互独立.

这里第一行有 $\binom{n}{2}$ 个式子，第二行有 $\binom{n}{3}$ 个式子，等等，因此共应满足

$$\binom{n}{2}+\binom{n}{3}+\cdots+\binom{n}{n}=2^n-n-1$$

个等式. 由三个事件的场合可看出同时满足这些关系式是必须的.

显然若 n 个事件相互独立，则它们中的任何 $m(2\leqslant m<n)$ 个事件也相互独立. 此外对于多个相互独立的事件也成立着类似于上述推论 1 及推论 2 的结果，读者试自行叙述并验证之.

三、事件独立性与概率的计算

从事件独立性的定义立刻能看出：若事件是独立的，则许多概率的计算就可以大为简化. 下面先举两个例子.

[相互独立事件至少发生其一的概率的计算] 若 $A_1, A_2, \cdots, A_n$ 是 n 个相互独立的事件，则由于

$$\overline{A_1\cup A_2\cup\cdots\cup A_n}=\bar{A}_1\bar{A}_2\cdots\bar{A}_n$$

因此

$$\begin{aligned} P(A_1\cup A_2\cup\cdots\cup A_n)&=1-P(\bar{A}_1\bar{A}_2\cdots\bar{A}_n) \\ &=1-P(\bar{A}_1)P(\bar{A}_2)\cdots P(\bar{A}_n) \end{aligned} \tag{6}$$

这个公式比起不独立的场合，要简便得多，它经常被用到.

[例 5] 假若每个人血清中含有肝炎病毒的概率为0.4%，混和100个人的血清，求此血清中含有肝炎病毒的概率.

[解] 以$A_i(i=1,2,\cdots,100)$记第i个人的血清含有肝炎病毒这一事件，可以认为它们相互独立，所求概率为$P(A_1\cup\cdots\cup A_{100})$，由(6)得

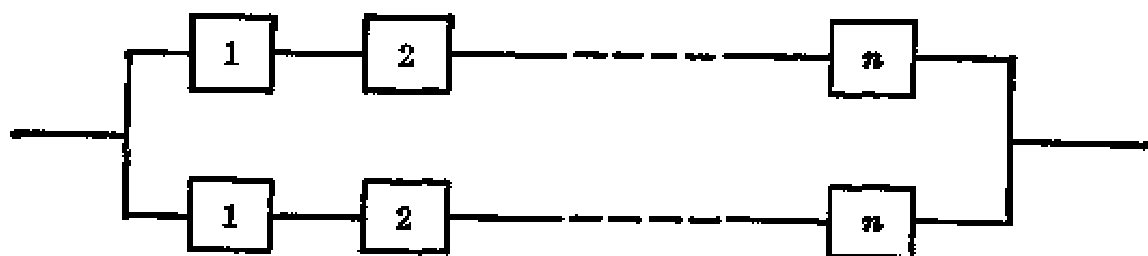
$$P(A_1\cup\cdots\cup A_{100})=1-P(\bar{A}_1)\cdots P(\bar{A}_{100})=1-0.996^{100}\approx 0.33$$

虽然每个人有病毒的概率很小，但是混和后则有很大的概率，在实际工作中，这类效应值得充分重视.

[在可靠性理论中的应用] 对于一个元件，它能正常工作的概率p，称为它的可靠性. 元件组成系统，系统正常工作的概率称为该系统的可靠性. 随着近代电子技术的迅猛发展，关于元件和系统可靠性的研究已发展成为一门新的学科——可靠性理论.

这里，我们通过一些例子来说明有关的概念.

[例 6] 如果构成系统的每个元件的可靠性均为r，$0<r<1$，且各元件能否正常工作是相互独立的，试求下面(附加通路)系统的可靠性.



[解] 每条通路要能正常工作，当且仅当该通路上各元件正常工作，故其可靠性为

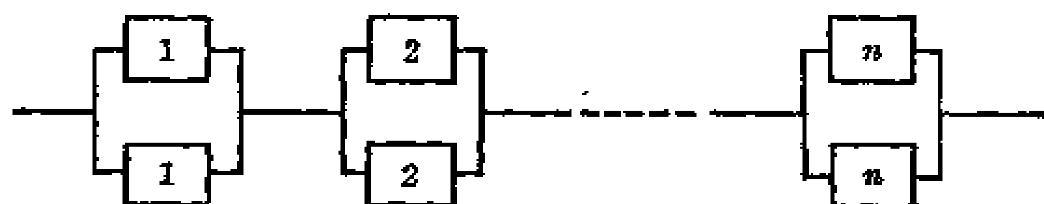
$$R_c=r^n$$

也即通路发生故障的概率为$1-r^n$，由于系统是由两通路并联而成的，两通路同时发生故障的概率为$(1-r^n)^2$，因此上述系统的可靠性为

$$R_S=1-(1-r^n)^2=r^n(2-r^n)=R_c(2-R_c)$$

注意到 $R_c<1$, 故 $R_S>R_c$, 所以附加通路能使系统的可靠性增加.

[例 7] 在前例条件下, 试求下面(附加元件)系统的可靠性.



[解] 每对并联元件的可靠性为

$$R'=1-(1-r)^2=r(2-r)$$

系统由每对并联元件串联而成, 故其可靠性为

$$R'_S=(R')^n=r^n(2-r)^n$$

显然 $R'_S>R_c$. 因此用附加元件的方法同样也能增加系统的可靠性.

利用数学归纳法不难证明当 $n\geqslant 2$ 时, $(2-r)^n>2-r^n$, 即 $R'_S>R_S$. 因此虽然上面两个系统同样由 $2n$ 个元件构成, 作用也相同, 但是第二种构成方式比第一种方式可靠性来得大. 寻找可靠性较大的构成方式也是可靠性理论的研究课题之一.

上述讨论可以看出, 元件与系统的可靠性是用概率来定义的, 所以概率论是研究可靠性理论的重要工具.

四、试验的独立性

有了事件独立性的概念, 我们可以定义试验的独立性. 直观上, 试验 E_1 与试验 E_2 要能说是独立的, 应是指试验 E_1 的结果的发生与试验 E_2 的结果的发生是独立的, 所以自然想到通过各试验的事件间的独立性来定义试验的独立性. 为了做到这点, 首先要构造一个能描述这些试验的公共的样本空间.

设试验 E_1 的样本空间是 $\Omega_1=\{\omega^{(1)}\}$, 试验 E_2 的样本空间是 $\Omega_2=\{\omega^{(2)}\}$, ……, 试验 E_n 的样本空间是 $\Omega_n=\{\omega^{(n)}\}$, 为描述这 n

次试验，应构造复合试验 E，它表示依次进行试验 $E_1, E_2, \cdots, E_n$，其样本点为

$$\omega=(\omega^{(1)}, \omega^{(2)}, \cdots, \omega^{(n)}) \tag{7}$$

这个样本空间记作

$$\Omega=\Omega_1\times\Omega_2\times\cdots\times\Omega_n \tag{8}$$

［例 8］ 若试验 E_1 是掷一枚硬币，$\Omega_1=\{$正，反$\}$，试验 E_2 是从装有红白黑三球的袋子中摸出一球，$\Omega_2=\{$红，白，黑$\}$，则复合试验 E 表示先掷一枚硬币再摸一球，它相应的样本空间 $\Omega=\Omega_1\times\Omega_2$ 由下列 6 个样本点构成：(正，红)，(正，白)，(正，黑)，(反，红)，(反，白)，(反，黑)。

接着我们可以引进"与第 k 次试验有关的事件"的概念，这种事件发生与否仅与第 k 次试验的结果有关，即为了判断某一样本点是否属于这个事件，只须察看它的第 k 个分量．例如在例 8 中，"第 2 次试验摸出红球"这一事件就是一个与第 2 次试验有关的事件．值得指出，必然事件 Ω 及不可能事件 $\varnothing$ 可以认为与所有的试验有关。因为 Ω 及 $\varnothing$ 的出现与否跟试验的任何结果没有关系．

现在若以 $\mathscr{A}_k$ 记与第 k 次试验有关的事件全体，则可以通过下列方式定义试验的独立性．

定义 若对于任意的

$$A^{(1)}\in\mathscr{A}_1,\ A^{(2)}\in\mathscr{A}_2,\ \cdots,\ A^{(n)}\in\mathscr{A}_n$$

均成立

$$P(A^{(1)}A^{(2)}\cdots A^{(n)})=P(A^{(1)})P(A^{(2)})\cdots P(A^{(n)}) \tag{9}$$

则称试验 $E_1, E_2, \cdots, E_n$ 是相互独立的．

注意到 $\Omega\in\mathscr{A}_i$，$i=1, 2, \cdots, n$，因此由定义立刻推出，若 n 个试验相互独立，则其中的 $m(2\leqslant m<n)$ 个试验也是相互独立的．

在例 8 中，若对样本空间 Ω 中的 6 个样本点都给定概率 $\frac{1}{6}$，则

不难验证试验 E_1 与试验 E_2 是相互独立的. 读者可以思考一下,是否还可以给定其他概率使试验 E_1 与试验 E_2 独立?

n 次有放回摸球所构成的 n 个试验是相互独立的,而 n 次不放回摸球模型则是 n 个试验不独立的简单例子.

特别重要的一类试验是所谓**重复独立试验**,这时 $\Omega_1=\Omega_2=\cdots=\Omega_n$,有关事件的概率保持不变,而且各次试验是相互独立的. 投 n 次硬币或进行 n 次有放回摸球是重复独立试验的简单例子. 重复独立试验是作为"在同样条件下重复试验"的数学模型而出现的,它在概率论中很有地位:因为随机现象的统计规律性只有在大量重复试验中才会显示出来.

下一节我们将研究一类最简单的重复独立试验——贝努里试验.

§3. 贝努里试验与直线上的随机游动

一、贝努里概型

在许多问题中,我们对试验感兴趣的是试验中某事件 A 是否发生. 例如在产品抽样检查中注意的是抽到废品,还是抽到好品;在掷硬币时注意的是出正面还是出反面. 在这类问题中我们可以把事件域取为 $\mathscr{F}=\{\varnothing, A, \bar{A}, \Omega\}$,并称出现 A 为"成功",出现 $\bar{A}$ 为"失败". 这种只有两个可能结果的试验称为**贝努里试验**.

有些试验的结果不止两个,譬如,在电报传输中,既要传送字母 $A, B, \cdots, Z$ 等,又要传送其他符号. 但是假如我们所关心的只是字母在传送中所占的百分比,而不再区别到底是哪一个字母,则我们可以把出现字母当作是成功,出现其他符号一律当作是失败,这时就可以把问题看作贝努里试验. 同样的,电子管的寿命可以是不小于 0 的任一数值. 但是有时根据需要,我们可以把寿命大于 500 小时的电子管当作合格品,其余都作为次品. 那么,这类问

题还是可以归结为贝努里试验. 这种例子可以举出不少.

在贝努里试验中, 首先是要给出下面概率

$$P(A)=p, \quad P(\bar{A})=q \tag{1}$$

显然 $p\geqslant 0$, $q\geqslant 0$, 且 $p+q=1$.

现在考虑重复进行 n 次独立的贝努里试验, 这里的"重复", 是指在每次试验中事件 A, 从而事件 $\bar{A}$ 出现的概率都保持不变. 这种试验称为 **n 重贝努里试验**, 记作 E^n.

n 重贝努里试验 E^n 的样本点形如:

$$(\omega_1, \omega_2, \cdots, \omega_n) \tag{2}$$

其中 ω_i 是 A 或 $\bar{A}$, 分别表示第 i 次试验中出现 A 或 $\bar{A}$, 显然这种样本点共有 2^n 个, 这是一个有限样本空间.

象对一般有限样本空间一样, 可以把样本点的任意子集作为事件. 利用试验的独立性, 只要通过(1)给定了 p, 那么就能给出所有样本点的概率, 从而也就给定了所有事件的概率. 例如 $(A, A, \cdots, A, \bar{A})$ 表示前 $n-1$ 次试验均出现事件 A, 而第 n 次试验出现事件 $\bar{A}$, 这是一个样本点, 其概率为 $p^{n-1}q$, 一般, 若一个样本点中 A 出现了 k 次, 因而 $\bar{A}$ 出现了 $n-k$ 次, 则其概率为 p^kq^{n-k}.

有时也需要考虑**可列重贝努里试验** E^∞ 的场合, 这时样本点形如

$$(\omega_1, \omega_2, \cdots, \omega_n, \cdots) \tag{3}$$

其中 ω_i 仍取 A 或 $\bar{A}$, 这时样本空间不再有限, 甚至也不可列, 事实上它可与 $[0, 1]$ 区间进行一一对应. 对这种试验就不能把样本空间的任意子集都看作事件.

贝努里试验是一种非常重要的概率模型, 它是"在同样条件下进行重复试验或观察"的一种数学模型; 特别在讨论某事件出现的频率时常用这种模型. 历史上, 贝努里概型是概率论中最早研究的模型之一, 也是得到最多研究的模型之一, 在理论上具有重

要意义；另一方面，它有着广泛的实际应用，例如在工业产品质量检查中，在群体遗传学中它都占有重要地位。在我们这门课程中，一些较为深入的结果也是结合贝努里概型进行讨论的.

二、贝努里概型中的一些分布

下面我们计算贝努里概型中所出现的一些事件的概率，这些概率非常重要.

1. 贝努里分布

若只进行一次贝努里试验，则或是事件 A 出现，或是事件 $\bar{A}$ 出现，其概率由(1)给出，称为**贝努里分布**，这是最简单的情况.

2. 二项分布

我们来确定 n 重贝努里试验中事件 A 出现 k 次的概率，这概率我们记之为 $b(k;n,p)$.

若以 B_k 记 n 重贝努里试验中事件 A 正好出现 k 次这一事件，而以 A_i 表示第 i 次试验中出现事件 A，以 $\bar{A}_i$ 表示第 i 次试验中出现 $\bar{A}$，则

$$B_k=A_1A_2\cdots A_k\bar{A}_{k+1}\cdots\bar{A}_n+\cdots+\bar{A}_1\bar{A}_2\cdots\bar{A}_{n-k}A_{n-k+1}\cdots A_n \quad (4)$$

右边的每一项表示在某 k 次试验中出现事件 A，在另外 $n-k$ 次试验中出现 $\bar{A}$，这种项共有 $\binom{n}{k}$ 个，而且两两互不相容. 由于试验的独立性

$$P(A_1A_2\cdots A_k\bar{A}_{k+1}\cdots\bar{A}_n)=P(A_1)P(A_2)\cdots P(A_k)P(\bar{A}_{k+1})\cdots P(\bar{A}_n)=p^kq^{n-k} \quad (5)$$

同理可得(4)中右边各项所对应事件的概率均为 p^kq^{n-k}，利用概率的加法定理知

$$P(B_k)=\binom{n}{k}p^kq^{n-k}$$

即

$$b(k;n,p)=\binom{n}{k}p^k q^{n-k},\quad k=0,1,2,\cdots,n \tag{6}$$

注意到 $b(k;n,p)$, $k=0,1,2,\cdots,n$, 是二项式$(q+ps)^n$展开式中 s^k 项的系数，因此(6)称为**二项分布**. 特别地

$$\sum_{k=0}^{n}b(k;n,p)=\sum_{k=0}^{n}\binom{n}{k}p^k q^{n-k}=(q+p)^n=1 \tag{7}$$

[例 1]　若在N件产品中有 M 件废品，现进行 n 次有放回的抽样检查，问共抽得 k 件废品的概率是多少？

[解]　由于抽样是有放回的，因此这是 n 重贝努里试验，若以 A 记各次试验中出现废品这一事件，则

$$p=P(A)=\frac{M}{N}$$

因此所求的概率为

$$b\left(k;n,\frac{M}{N}\right)=\binom{n}{k}\left(\frac{M}{N}\right)^k\left(1-\frac{M}{N}\right)^{n-k}$$

这概率在第一章 § 3 中曾出现过.

*[例 2]　在群体遗传学中，假定可遗传的指标是依赖于基因的. 基因总是成对出现并且具有两种型式 A 及 a 中的一种. 假定每一代具有 $2N$ 个基因，则其中 A 所占的比例数$\frac{i}{2N}$称为**基因频率**. 基因频率的变化过程对于群体的进化情况有着密切的关系.

如果进行的是随机交配，也就是说任何个体有同样的机会和任何其他个体配种，则遗传学中对基因的遗传作下面假定：子代个体是按贝努里概型从上一代每个亲体中取得基因的.

因此若上一代的基因频率为$\frac{i}{2N}$，则下一代的基因频率为$\frac{j}{2N}$的概率由二项分布给出

$$p_j=\binom{2N}{j}\left(\frac{i}{2N}\right)^j\left(1-\frac{i}{2N}\right)^{2N-j}$$

在这个例子中基因频率 $\frac{i}{2N}$ 相当于贝努里试验中出现成功的概率 p.

3. 几何分布

现在讨论在贝努里试验中首次成功出现在第 k 次试验的概率,要使首次成功出现在第 k 次试验,必须而且只须在前 $k-1$ 次试验中都出现事件 $\bar{A}$,而第 k 次试验出现 A,因此若沿用推导二项分布时所用的记号 A_i 及 $\bar{A}_i$,则这事件(记为 W_k)可表为

$$W_k = \bar{A}_1\bar{A}_2\cdots\bar{A}_{k-1}A_k \tag{8}$$

利用事件的独立性,其概率为

$$P(W_k) = P(\bar{A}_1)P(\bar{A}_2)\cdots P(\bar{A}_{k-1})P(A_k) = q^{k-1}p$$

记

$$g(k;p) = q^{k-1}p,\ \ k=1,2,\cdots \tag{9}$$

$g(k;p)$ 是几何级数的一般项,因此(9)称为几何分布. 特别地

$$\sum_{k=1}^{\infty} g(k;p) = \sum_{k=1}^{\infty} q^{k-1}p = p\frac{1}{1-q} = 1 \tag{10}$$

几何分布给出了等待事件 A 出现共等了 k 次的概率,这类概率在许多问题中出现,我们在第一章 §3 中曾遇到它. 下面是一个模型化了的例子.

[例 3] 一个人要开门,他共有 n 把钥匙,其中仅有一把是能开这门的. 他随机地选取一把钥匙开门,即在每次试开时每一把钥匙都以概率 $\frac{1}{n}$ 被使用,这人在第 s 次试开时成功的概率是多少?

[解] 这是一个贝努里试验, $p=\frac{1}{n}$,由(9),所求概率为

$$g\left(s;\frac{1}{n}\right) = \left(\frac{n-1}{n}\right)^{s-1}\frac{1}{n}$$

应当指出，讨论事件A的首次出现事实上牵涉到可列次贝努里试验．因此它的样本空间应取为(3)所表示的样本点全体，这个空间不是可列的．因此不能把它的一切子集都作为事件．在上述讨论中，我们只是把用(8)表出的W_k作为事件，并以(9)给出其概率，这里的W_k事实上也包含了不可列个用(3)表出的样本点．当然在讨论这个问题时，我们也可以干脆把W_k作为样本点，这时样本空间取为$\{W_1, W_2, W_3, \cdots\}$，并以(9)给出其概率，这就成了一个离散样本空间．

下面讨论的是更复杂一点的情况，即巴斯卡(Pascal)分布，它可以看作几何分布的一种推广．

4. 巴斯卡分布

考虑相继的贝努里试验，让我们考察要多长时间才会出现第r次成功．

若第r次成功发生在第ν次试验，则必然有$\nu \geqslant r$．

我们以C_k表示第r次成功发生在第k次试验这一事件，并以$f(k;r,p)$记其概率，要C_k发生当且仅当前面的$k-1$次试验中有$r-1$次成功，$k-r$次失败，而第k次试验的结果为成功，这些事件的概率分别为$\binom{k-1}{r-1}p^{r-1}q^{k-r}$与$p$，于是利用试验的独立性，得到

$$P(C_k)=\binom{k-1}{r-1}p^{r-1}q^{k-r}\cdot p=\binom{k-1}{r-1}p^r q^{k-r}$$

即

$$f(k;r,p)=\binom{k-1}{r-1}p^r q^{k-r},\quad k=r,\ r+1,\ \cdots \tag{11}$$

注意到

$$\sum_{k=r}^{\infty}f(k;r,p)=\sum_{k=r}^{\infty}\binom{k-1}{r-1}p^r q^{k-r}$$

$$=\sum_{l=0}^{\infty}\binom{r+l-1}{r-1}p^r q^l=\sum_{l=0}^{\infty}\binom{-r}{l}(-1)^l p^r q^l$$

$$=p^r(1-q)^{-r}=1 \tag{12}$$

这里利用了推广的二项系数公式$\binom{-r}{l}=(-1)^l\binom{r+l-1}{l}$.

$f(k;r,p)$称为**巴斯卡分布**，特别当$r=1$时，我们得到几何分布．巴斯卡是十七世纪的法国数学家，他对概率论的早期发展作出过重大贡献，巴斯卡分布是他在解决某一问题时导出的．

［例4］ 两个同类型的系统，开始时它们各有N个备件，每当出现故障时，就调换一个备件．假定两个系统的运行条件相同，那么当一个系统需要备件而备件已用光时，另一个系统可以还剩有r个备件($r=0,1,2,\cdots,N$)，我们分别记对应的概率为u_r，试求u_r的表达式．

［解］ 因为两系统类型及运行条件都相同，若只考察出故障的时刻，并把第一个系统出现故障看作出现事件A，而第二个系统出故障看作出现事件$\bar{A}$，则可把故障的出现看作贝努里试验，相应的概率$p=\frac{1}{2}$.

要第一个系统缺备件而第二个系统剩r件，应该是A出现$N+1$次(前N次用去所有N个备件，最后一次故障发生时缺乏调换的备件)而$\bar{A}$出现$N-r$次，这事件的概率为

$$f\left(2N-r+1;N+1,\frac{1}{2}\right)=\binom{2N-r}{N}\left(\frac{1}{2}\right)^{2N-r+1}$$

对于第二个系统先缺备件的情况可同样考虑，因此所求的概率为

$$u_r=2\cdot f\left(2N-r+1;N+1,\frac{1}{2}\right)=\binom{2N-r}{N}2^{-2N+r} \tag{13}$$

三、直线上的随机游动

考虑x轴上的一个质点，假定它只能位于整数点，在时刻$t=0$

时，它处于初始位置 a(a 是整数)，以后每隔单位时间，它总受到一个外力的随机作用，使位置发生变化，分别以概率 p 及概率 $q=1-p$ 向正的或负的方向移动一个单位，我们所关心的是质点在时刻 $t=n$ 时的位置. 用这种方式描述的质点运动称为**随机游动**.

若质点可以在整个数轴的整数点上游动，则称这种随机游动为**无限制随机游动**. 若在某点 d 设有一个吸收壁，质点一到达这点即被吸收而不再游动，因而整个游动也就结束了，这种随机游动称为**在 d 点有吸收壁的随机游动**. 此外还可以考虑带有反射壁及弹性壁的随机游动. 在一个随机游动中还可以具有不只一个壁.

当 $p=q=\frac{1}{2}$ 时，随机游动称为**对称的**，这时质点向左或向右移动的可能性相等.

自然科学中的大量问题归结为随机游动问题，例如随机游动模型可以作为布朗运动的初步近似. 概率论中的一些古典问题也引导到随机游动问题，事实上，随机游动可以看作是贝努里试验的一种描述法.

关于随机游动，已进行过许多研究，我们只介绍它的两个最简单的模型.

无限制随机游动　假定质点在时刻 0 从原点出发，以 S_n 记它在时刻 $t=n$ 时的位置. 为了使质点在时刻 $t=n$ 时位于 k(k 也可以是负整数)，必须而且只须在前 n 次游动中向右游动的次数比向左游动的次数多 k 次，若以 x 记它在前 n 次游动中向右游动的次数，y 记向左移动的次数，则

$$\begin{cases} x+y=n \\ x-y=k \end{cases} \tag{14}$$

即 $x=\frac{n-k}{2}$，因为 x 是整数，因此 k 必须与 n 具有相同的奇偶性.

事件$\{S_n=k\}$发生相当于要求在前n次游动中有$\frac{n+k}{2}$次向右，$\frac{n-k}{2}$次向左，利用二项分布即得

$$P\{S_n=k\}=\binom{n}{\frac{n+k}{2}}q^{\frac{n-k}{2}}p^{\frac{n+k}{2}} \tag{15}$$

当k与n奇偶性相反时，概率为0.

两端带有吸收壁的随机游动 假定质点在时刻$t=0$时，位于$x=a$，而在$x=0$及$x=a+b$处各有一个吸收壁，我们来求质点在$x=0$被吸收及在$x=a+b$被吸收的概率．用的是差分方程法.

若以p_n记质点的初始位置为n而随机游动被在0点的吸收壁吸收的概率，以q_n记初始位置为n但在$a+b$点被吸收的概率．显然

$$\begin{aligned} &p_{a+b}=0, \qquad q_{a+b}=1, \\ &p_0=1, \qquad q_0=0 \end{aligned} \tag{16}$$

如果某时刻质点位于$x=n$，则它要被$x=0$吸收有两种方式来实现：一种是接下去一次移动是向右的而最后被$x=0$吸收；另一种是接下去一次移动是向左的而最后被$x=0$吸收．所以按全概率公式有

$$p_n=p\cdot p_{n+1}+q\cdot p_{n-1}, \quad n=1,2,\cdots,a+b-1 \tag{17}$$

这样，我们得到了关于p_n的一个有限差分方程(17)，再利用边界条件(16)就可以求得p_n，$n=1,2,\cdots,a+b-1$．下面我们来求解.

把(17)改写为

$$\begin{gathered} q(p_n-p_{n-1})=p(p_{n+1}-p_n) \\ n=1,2,\cdots,a+b-1 \end{gathered} \tag{18}$$

我们先就$p=q=\frac{1}{2}$来解上面方程，这相应于对称随机游动的场合．这时方程成为

$$p_n - p_{n-1} = p_{n+1} - p_n$$

若设

$$p_{n+1} - p_n = p_n - p_{n-1} = \cdots\cdots = p_1 - p_0 = c$$

这里 c 是常数．由此可得

$$p_n = p_0 + nc$$

因为 $p_0 = 1$, $p_{a+b} = 0$, 故有

$$p_n = 1 - \frac{n}{a+b} \tag{19}$$

因此从 $x=a$ 出发而被 $x=0$ 吸收的概率为

$$p_a = 1 - \frac{a}{a+b} = \frac{b}{a+b} \tag{20}$$

同样的方法可以求得在对称随机游动场合，从 $x=a$ 出发而被 $x=a+b$ 吸收的概率为

$$q_a = \frac{a}{a+b} \tag{21}$$

注意到 $p_a + q_a = 1$, 即质点最终一定要被吸收．

在一般场合，即 $p \neq q$ 时，由(18)得

$$q^n \prod_{k=1}^{n} (p_k - p_{k-1}) = p^n \prod_{k=1}^{n} (p_{k+1} - p_k)$$

因此

$$q^n(p_1 - p_0) = p^n(p_{n+1} - p_n), \quad n = 0, 1, 2, \cdots, a+b-1$$

由(16), $p_0 = 1$, 所以

$$p_{n+1} - p_n = \left(\frac{q}{p}\right)^n (p_1 - 1), \quad n = 0, 1, 2, \cdots, a+b-1 \tag{22}$$

因为

$$p_{a+b} - p_n = \sum_{k=n}^{a+b-1} (p_{k+1} - p_k) = \sum_{k=n}^{a+b-1} \left(\frac{q}{p}\right)^k (p_1 - 1)$$

$$=(p_1-1)\frac{\left(\frac{q}{p}\right)^n-\left(\frac{q}{p}\right)^{a+b}}{1-\frac{q}{p}}$$

由(16)，$p_{a+b}=0$，故

$$p_n=(1-p_1)\frac{\left(\frac{q}{p}\right)^n-\left(\frac{q}{p}\right)^{a+b}}{1-\frac{q}{p}} \tag{23}$$

因 $p_0=1$，故

$$1=(1-p_1)\frac{1-\left(\frac{q}{p}\right)^{a+b}}{1-\frac{q}{p}} \tag{24}$$

由(23)及(24)立即得到

$$p_n=\frac{\left(\frac{q}{p}\right)^n-\left(\frac{q}{p}\right)^{a+b}}{1-\left(\frac{q}{p}\right)^{a+b}} \tag{25}$$

因此质点从 $x=a$ 出发，在 $x=0$ 被吸收的概率为

$$p_a=\frac{\left(\frac{q}{p}\right)^a-\left(\frac{q}{p}\right)^{a+b}}{1-\left(\frac{q}{p}\right)^{a+b}}=\frac{1-\left(\frac{p}{q}\right)^b}{1-\left(\frac{p}{q}\right)^{a+b}} \tag{26}$$

用同样的方法可以求得从 $x=a$ 出发，而在 $x=a+b$ 被吸收的概率为

$$q_a=\frac{1-\left(\frac{q}{p}\right)^a}{1-\left(\frac{q}{p}\right)^{a+b}} \tag{27}$$

注意到 $p_a+q_a=1$，因此质点最后一定要被吸收。

另外，就是表达式(26)及(27)当 $p=q=\frac{1}{2}$ 时没有意义，这时

解为(20)及(21)，但是(20)及(21)也可从(26)及(27)中令 $p\to q$ 并用洛必达法则得到.

四、推广的贝努里试验与多项分布

二项分布可以容易地推广到 n 次重复独立试验且每次试验可能有若干个结果的情形. 把每次试验的可能结果记为 $A_1, A_2, \cdots, A_r$，而 $P(A_i)=p_i, i=1, 2, \cdots, r$，且

$$p_1+p_2+\cdots+p_r=1, \quad p_i\geqslant 0 \tag{28}$$

当 $r=2$ 时，我们得到贝努里试验.

在这种推广的贝努里试验中，不难导出，在 n 次试验中 A_1 出现 k_1 次，A_2 出现 k_2 次，…，A_r 出现 k_r 次的概率为

$$\frac{n!}{k_1!k_2!\cdots k_r!}p_1^{k_1}p_2^{k_2}\cdots p_r^{k_r} \tag{29}$$

这里 $k_i\geqslant 0$，且 $k_1+k_2+\cdots+k_r=n$.

公式(29)称为**多项分布**，因为它是 $(p_1+p_2+\cdots+p_r)^n$ 的展开式的一般项，而且由(28)知

$$\sum_{\substack{k_i\geqslant 0\\ k_1+k_2+\cdots+k_r=n}}\frac{n!}{k_1!k_2!\cdots k_r!}p_1^{k_1}p_2^{k_2}\cdots p_r^{k_r}=1 \tag{30}$$

显然多项分布是二项分布的推广. 二项分布中的许多结果都能平行地推广到多项分布的场合，以后我们只详细讨论二项分布的有关问题.

在产品检查中，若对产品质量所用的标准不只是好品与废品，而是分得更细，例如有一等品，二等品，三等品，等外品四类，则从中取出 n 件，求一等品有 k_1 件，二等品有 k_2 件，三等品有 k_3 件，等外品有 k_4 件的概率时便得到多项分布.

［例 5］ 人类的血型分为 O, A, B, AB 四型，假定某地区的居民中这四种血型的人的百分比分别为 0.4, 0.3, 0.25, 0.05，若从此地区居民中随机地选出 5 人，求有两个为 O 型，其他三个分别是 A,

B, AB 型的概率。

[解] 推广的贝努里试验可以用于这个场合，所求的概率为

$$P=\frac{5!}{2!1!1!1!}\times 0.4^2\times 0.3\times 0.25\times 0.05=0.036$$

也可以研究平面上或空间的随机游动，下面是简单的一例。

[例 6] （平面上随机游动）一质点从平面上某点出发，等可能地向上，下，左，右方向移动，每次移动的距离为 1，求经过 $2n$ 次移动后回到出发点的概率。

[解] 这可以归结为上述推广的贝努里试验的问题，分别以事件 A_1, A_2, A_3, A_4 表示质点向上，下，左，右移动一格，则 $p_1=p_2=p_3=p_4=\frac{1}{4}$. 若要在 $2n$ 次移动后回到原来的出发点，则向左移动的次数与向右移动的次数应该相等，向上移动的次数与向下移动的次数也应该相等．而总移动次数为 $2n$，因此所求概率为

$$P=\sum_{k+m=n}\frac{(2n)!}{(k!)^2(m!)^2}\left(\frac{1}{4}\right)^{2n}=\sum_{k=0}^{n}\frac{(2n)!}{(k!)^2[(n-k)!]^2}\left(\frac{1}{4}\right)^{2n}$$

$$=\left(\frac{1}{4}\right)^{2n}\frac{(2n)!}{(n!)^2}\sum_{k=0}^{n}\left[\frac{n!}{k!(n-k)!}\right]^2$$

$$=\left(\frac{1}{4}\right)^{2n}\binom{2n}{n}\sum_{k=0}^{n}\binom{n}{k}^2=\left(\frac{1}{4}\right)^{2n}\binom{2n}{n}^2 \tag{31}$$

最后一个等式用到第一章 § 3 中的公式。

§ 4. 二项分布与普阿松分布

一、二项分布的性质及计算

1. 二项分布的计算

在上一节中我们导出了在 n 次贝努里试验中正好出现 k 次成功的概率 $b(k; n, p)$：

$$b(k;n,p)=\binom{n}{k}p^k q^{n-k}, \quad k=0,1,2,\cdots,n \tag{1}$$

其中 $q=1-p$.

$b(k;n,p)$, $k=0,1,2,\cdots,n$ 称为二项分布，在概率论中占有很重要的地位. 由于在许多实际问题中出现二项分布，并且要计算其数值，因此讨论二项分布的计算显得非常重要.

二项分布有现成的表可查，这种表对不同的 n 及 p 给出了 $b(k;n,p)$ 的数值.

为了增加对二项分布的感性认识及计算的需要，我们选取了下列二项分布数值表.

表 1 中给出了对于 $n=20$ 及 $p_1=0.1$, $p_2=0.3$, $p_3=0.5$ 的二项分布数值表.

二项分布表只对 $p\leqslant 0.5$ 给出，因为对于 $p>0.5$ 的概率不难经下式计算得到.

$$b(k;n,p)=b(n-k;n,1-p) \tag{2}$$

表 1 二项分布数值表

k	$b(k;20,p)$			k	$b(k;20,p)$		
	$p_1=0.1$	$p_2=0.3$	$p_3=0.5$		$p_1=0.1$	$p_2=0.3$	$p_3=0.5$
0	0.1216	0.0008	—	11	—	0.0120	0.1602
1	0.2702	0.0068	—	12	—	0.0039	0.1201
2	0.2852	0.0278	0.0002	13	—	0.0010	0.0739
3	0.1901	0.0716	0.0011	14	—	0.0002	0.0370
4	0.0898	0.1304	0.0046	15	—	—	0.0148
5	0.0319	0.1789	0.0148	16	—	—	0.0046
6	0.0089	0.1916	0.0370	17	—	—	0.0011
7	0.0020	0.1643	0.0739	18	—	—	0.0002
8	0.0004	0.1144	0.1201	19	—	—	—
9	0.0001	0.0654	0.1602	20	—	—	—
10	—	0.0308	0.1762				

当 $p>0.5$ 时，$1-p<0.5$，仍能利用分布数值表.

为了对二项分布的变化情况有个直观了解，我们把表 1 中的几个分布用图 1 表示出来.

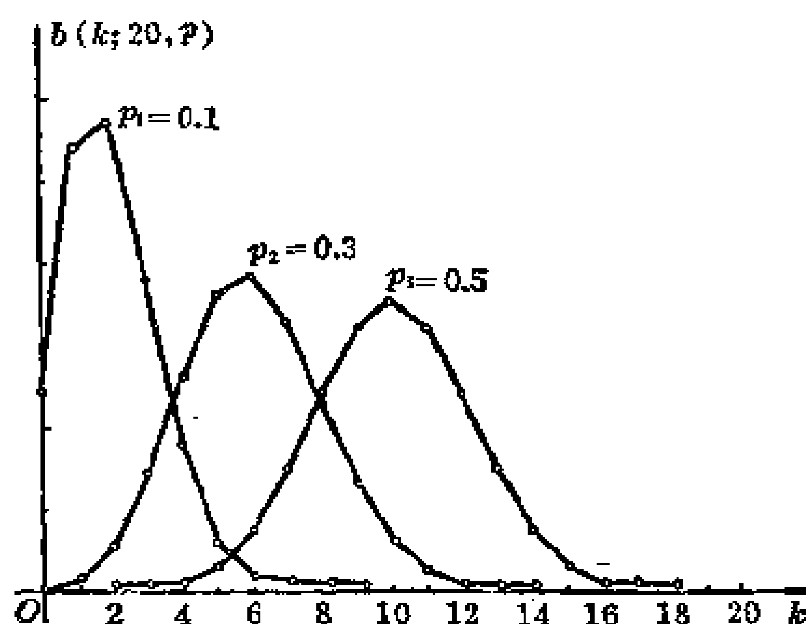


图 1　二项分布图

从图中可以看出，对于固定的 n 及 p，当 k 增加时，$b(k;n,p)$ 先随之增加并达到某极大值，以后又下降．此外，当概率 p 越与 $\frac{1}{2}$ 接近时，分布越接近对称.

［例 1］　一大批电子管中有 10% 已损坏，若我们从这批电子管中随机地选取 20 只来组成一个线路，问这线路能正常工作(即所选的 20 个电子管全部是好的)的概率是多少？

［解］　因为这批电子管数量很大，因此可近似地把选取 20 个管作为进行独立试验．这是一个贝努里概型问题，若把选得好电子管作为成功．其概率为 0.9，则所求的概率为

$$b(20;20,0.9)=\binom{20}{20}0.9^{20}=0.9^{20}$$

这个数值可以利用对数表计算，但更方便的是利用表 1 计算，由(2)

$$b(20;\ 20,\ 0.9)=b(0;\ 20,\ 0.1)$$

利用表 1 知所求的概率为 0.1216.

[例 2] （血清的试验）设在家畜中感染某种疾病的概率是 30%，新发现了一种血清可能对预防此病有效，为此对 20 只健康的动物注射这种血清．若注射后只有一只动物受感染，我们应对此种血清的作用作何评价？

假如血清毫无价值，那么注射后的动物受感染的概率还是 30%，则这 20 只动物中有 k 只受感染的概率为 $b(k;\ 20,\ 0.3)$.

发生只有一只动物受感染或更好的情况（无动物受感染）的概率为

$$b(0;\ 20,\ 0.3)+b(1;\ 20,\ 0.3)=0.0008+0.0068=0.0076$$

这个概率相当小，因此我们不能认为血清毫无价值.

如果注射后的 20 只动物中有 4 只受感染，我们是否相信此种血清有效？这个问题留给读者思考.

2. 二项分布的性质

我们来考察 $b(k;\ n,\ p)$ 随 k 及 n 变化的情况．从图 1 可以看出，当 n 固定时，$b(k;\ n,\ p)$ 先随 k 增加而增大，达到某一极大值后又逐渐下降．现在对它进行严格讨论.

由于对 $0<p<1$,

$$\frac{b(k;n,p)}{b(k-1;n,p)}=\frac{(n-k+1)p}{kq}=1+\frac{(n+1)p-k}{kq} \tag{3}$$

因此

$$\begin{aligned}
&\text{当 } k<(n+1)p \text{ 时}, b(k;n,p)>b(k-1;n,p)\\
&\text{当 } k=(n+1)p \text{ 时}, b(k;n,p)=b(k-1;n,p)\\
&\text{当 } k>(n+1)p \text{ 时}, b(k;n,p)<b(k-1;n,p)
\end{aligned} \tag{4}$$

因为 $(n+1)p$ 不一定是正整数，而二项分布中的 k 只取正整数值，所以存在正整数 m，使得 $(n+1)p-1<m\leqslant(n+1)p$，而且

当 k 从 0 变到 n 时，$b(k;n,p)$ 起先单调上升，当 $k=m$ 时达到极大值，后来又单调下降．但若 $(n+1)p=m$ 则这时 $b(m;n,p)=b(m-1;n,p)$ 同时达到极大值．

使 $b(k;n,p)$ 取最大值的项 $b(m;n,p)$ 称为 $b(k;n,p)$ 的中心项，而 m 称为最可能成功次数．由上面讨论知 $m=[(n+1)p]$（即 m 是 $(n+1)p$ 的整数部分）．若 $(n+1)p$ 是整数，则 $m-1$ 亦为最可能成功次数．

［例 3］ 设每颗子弹打中飞机的概率为 0.01，问在 500 发中打中飞机的最可能次数是多少？并求其相应的概率．

［解］ 这是贝努里概型，打中飞机子弹数服从二项分布．$n=500$，$p=0.01$，$(n+1)p=5.01$，$[(n+1)p]=5$．所以最可能成功次数为 5．相应概率为

$$b(5;500,0.01)=\binom{500}{5}(0.01)^5\cdot(0.99)^{495}=0.117635 \qquad (5)$$

应该注意，若 $0<p<1$，则当 n 值相当大时，即便是最可能成功次数 m 发生的概率也相当小，对于其他的 k，则 $b(k;n,p)$ 自然更小了，以后将会看到最可能成功次数 m 发生的概率接近于

$$(2\pi npq)^{-\frac{1}{2}}$$

（当 n 相当大时），因此当 $n\to\infty$ 时，这概率趋于 0．

3．应用的例子

我们举一些应用的例子，说明二项分布的重要性，同时也提出一些问题．

［例 4］ 保险事业是最早使用概率论的部门之一．保险公司为了估计企业的利润，需要计算各种各样的概率，下面是典型问题之一．若一年中某类保险者里面每个人死亡的概率等于 0.005，现有 10000 个这类人参加人寿保险，试求在未来一年中在这些保险

者里面，(1) 有 40 个人死亡的概率；(2) 死亡人数不超过 70 个的概率。

［解］ 作为初步近似，可以利用贝努里概型，$n=10000$，$p=0.005$，设 μ 为未来一年中这些人里面死亡的人数，则所求的概率分别为

$$(1)\ b(40;\ 10000,\ 0.005)=\binom{10000}{40}(0.005)^{40}(0.995)^{9960} \tag{6}$$

$$(2)\ P\{\mu\leqslant 70\}=\sum_{k=0}^{70}b(k;\ 10000,\ 0.005)=\sum_{k=0}^{70}\binom{10000}{k}(0.005)^{k}(0.995)^{10000-k} \tag{7}$$

直接计算这些数值相当困难，要有更好的计算方法。

［例 5］ 甲，乙两容器，容量各为 1 升，每个各含 2.7×10^{22} 个气体分子，现将两容器接触，经过相当长的时间后$\left(\text{即这时每个分子落在两容器中的概率各为}\frac{1}{2}\right)$，求两容器中分子数之差超过分子总数的 100 亿分之一的概率。

［解］ 分子总数为 5.4×10^{22}，其 100 亿分之一（即 10^{-10}）为 5.4×10^{12}，设混和后甲容器的分子数为 μ，则乙容器的分子数为 $5.4\times10^{22}-\mu$，现要求事件

$$|\mu-(5.4\times10^{22}-\mu)|>5.4\times10^{12}$$

即事件

$$|\mu-2.7\times10^{22}|>2.7\times10^{12} \tag{8}$$

的概率，由二项分布

$$P\{|\mu-2.7\times10^{22}|>2.7\times10^{12}\}$$

$$=\sum_{k} b\left(k;\ 5.4\times10^{22},\ \frac{1}{2}\right) \tag{9}$$

上式和号对一切满足$|k-2.7\times10^{22}|>2.7\times10^{12}$的$k$求和，这概率是无法直接计算的.

从上面两个例子中可以看到，计算二项分布的数值时，由于试验次数n经常很大，因此实际计算$b(k;n,p)$及$\sum_k b(k;n,p)$都很困难，有时甚至不可能．例如(9)中须算的项数有5.4×10^{12}之多，逐项计算是不可能的.

在这种情况下，寻找更有效的计算法是必要的，即便是近似公式也好．这可以利用概率论中的极限定理来实现，关于极限定理的讨论将在第五章进行.

二、二项分布的普阿松(poisson)逼近

在很多应用问题中，我们常常遇到这样的贝努里试验，其中，相对地说，n大，p小，而乘积$\lambda=np$大小适中．在这种情况下，有一个便于使用的近似公式，下面我们来推导它.

定理(普阿松)　在贝努里试验中，以p_n代表事件A在试验中出现的概率，它与试验总数n有关，如果$np_n\to\lambda$，则当$n\to\infty$时，

$$b(k;n,p_n)\to\frac{\lambda^k}{k!}e^{-\lambda} \tag{10}$$

[证明]　记$\lambda_n=np_n$，则

$$\begin{aligned}b(k;n,p_n)&=\binom{n}{k}p_n^k(1-p_n)^{n-k}\\&=\frac{n(n-1)\cdots(n-k+1)}{k!}\left(\frac{\lambda_n}{n}\right)^k\left(1-\frac{\lambda_n}{n}\right)^{n-k}\\&=\frac{\lambda_n^k}{k!}\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{k-1}{n}\right)\left(1-\frac{\lambda_n}{n}\right)^{n-k}\end{aligned}$$

由于对固定的k有

$$\lim_{n\to\infty}\lambda_n^k=\lambda^k,\quad \lim_{n\to\infty}\left(1-\frac{\lambda_n}{n}\right)^{n-k}=e^{-\lambda}$$

及

$$\lim_{n\to\infty}\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{k-1}{n}\right)=1$$

因此

$$\lim_{n\to\infty}b(k;n,p_n)=\frac{\lambda^k}{k!}e^{-\lambda}$$

定理证毕.

$p(k;\lambda)=\frac{\lambda^k}{k!}e^{-\lambda}$, $k=0,1,2,\cdots$称为普阿松分布，λ 是它的参数. 特别地

$$\sum_{k=0}^{\infty}p(k;\lambda)=\sum_{k=0}^{\infty}\frac{\lambda^k}{k!}e^{-\lambda}=e^{-\lambda}\cdot e^{\lambda}=1 \tag{11}$$

普阿松分布是概率论中很重要的一种分布.

在应用中，当 p 相当小(一般当 $p\leqslant 0.1$)时，我们用下面近似公式

$$b(k;n,p)\approx\frac{(np)^k}{k!}e^{-np} \tag{12}$$

[例 6] 假如生三胞胎的概率为 10^{-4}，求在 100000 次生育中，有 0, 1, 2 次生三胞胎的概率

[解] 这可以看作贝努里试验；$n=100000$, $p=0.0001$，所求的概率直接计算为

$$b(0;\ 100000,\ 0.0001)=0.000045378$$

$$b(1;\ 100000,\ 0.0001)=0.00045382$$

$$b(2;\ 100000,\ 0.0001)=0.0022693$$

这时也可用普阿松逼近，$\lambda=np=10$，而

$$p(0,10)=0.00004540$$

$$p(1,10)=0.0004540$$

$$p(2,10)=0.002270$$

可见近似程度很令人满意.

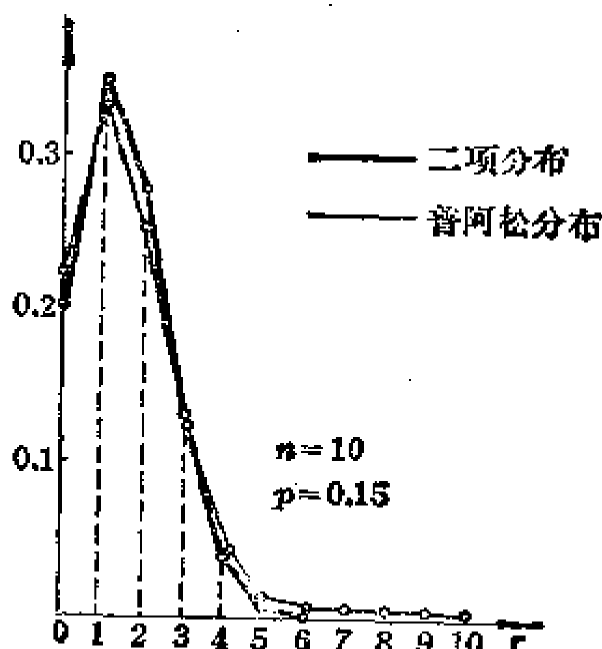


图 2 二项分布与普阿松分布

图 2 给出了二项分布逼近普阿松分布的一个图示，吻合程度甚好.

三、普阿松分布

在历史上普阿松分布是作为二项分布的近似，于 1837 年由法国数学家普阿松引入的. 近数十年来，普阿松分布日益显示其重要性，成了概率论中最重要的几个分布之一. 其原因主要是下面两点.

首先是已经发现许多随机现象服从普阿松分布. 这种情况特别集中在两个领域中. 一是社会生活，对服务的各种要求：诸如电话交换台中来到的呼唤数，公共汽车站来到的乘客数等等都近似地服从普阿松分布，因此在运筹学及管理科学中普阿松分布占有很突出的地位；另一领域是物理学. 放射性分裂落到某区域的质点数，热电子的发射，显微镜下落在某区域中的血球或微生物的数目等等都服从普阿松分布.

其次，对普阿松分布的深入研究（特别是通过随机过程的研究）已发现它具有许多特殊的性质和作用，打个不很恰当的譬喻，似乎普阿松分布是构造随机现象的“基本粒子”之一.

图 3 是对于不同 λ 值的普阿松分布图，为了计算普阿松分布的数值，有许多专门的表格可供查用.

下面提供两个有关的统计资料作为例子.

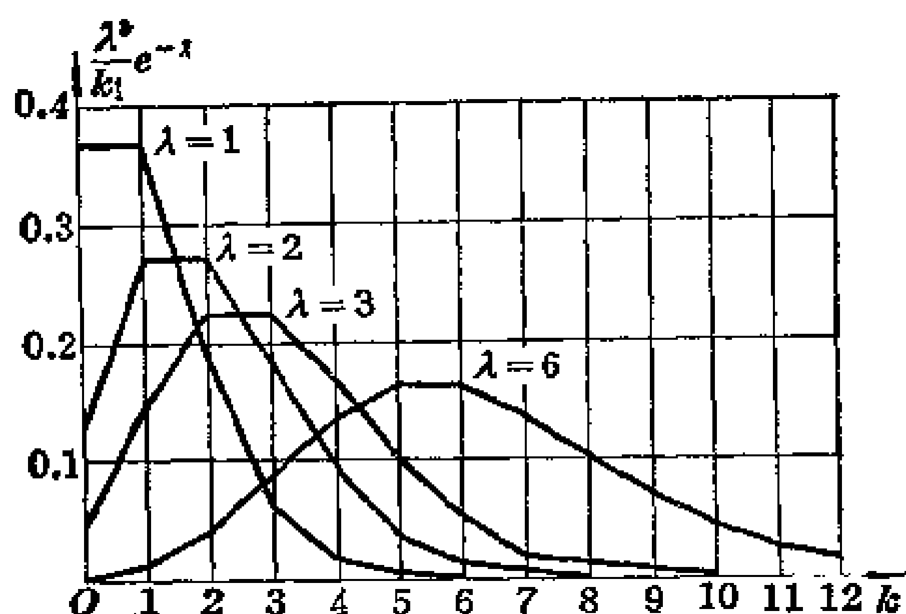


图 3 普阿松分布图

［例 7］ 对上海某公共汽车站的客流进行调查，统计了某天上午 10:30 至 11:47 左右每隔 20 秒钟来到的乘客批数(每批可能有数人同时来到)，共得 230 个记录，我们分别计算了来到 0 批，1 批，2 批，3 批，4 批及 4 批以上乘客的时间区间的频数，结果列于表 2 中，其相应的频率与 $\lambda=0.87$ 的普阿松分布符合得很好.

表 2 公共汽车客流统计

来到批数 i	0	1	2	3	≥4	总 共
频数 n_i	100	81	34	9	6	230
频率 $f_i=\frac{n_i}{n}$	0.43	0.35	0.15	0.04	0.03	
$p_i=\frac{\lambda^i}{i!}e^{-\lambda}$	0.42	0.36	0.16	0.05	0.01	

［例 8］ 放射性物质放射出的 α 质点数是服从普阿松分布的有名例子. 1910 年卢瑟福(Rutherford) 等人的著名实验揭露了这个事实.

在这个实验中，观察了长为 7.5 秒的时间间隔里到达某指定

区域的质点数，共观察了 $N=2608$ 次，表 3 给出观察值与理论值的对照，N_k 表示在 N 次观察中发生"在 7.5 秒内落到指定区域的质点数为 k"的观察次数，理论值是$NP(k;3.870)$，理论值与实验值很近似.

表 3　卢瑟福实验理论值与实验值对照表①

k	N_k	$NP(k;3.870)$
0	57	54.399
1	203	210.523
2	383	407.361
3	525	525.496
4	532	508.418
5	408	393.515
6	273	253.817
7	139	140.325
8	45	67.882
9	27	29.189
$k\geqslant 10$	16	17.075
总　计	2608	2608.000

在说明了普阿松分布的常见性之后，我们转入介绍产生普阿松分布的机制. 经过研究，已经弄清了服从普阿松分布的条件. 为了便于理解，我们将结合电话呼叫流来叙述这个重要结果.

我们先证明一个以后屡次要用到的数学分析结论.

引理　若 $f(x)$是连续函数(或单调函数)，且对一切 x,y(或一切 $x\geqslant 0, y\geqslant 0$)成立

$$f(x)f(y)=f(x+y) \tag{13}$$

则

$$f(x)=a^x \tag{14}$$

① 引自费勒，《概率论及其应用，离散空间》(上册)，科学出版社，1964，第 155 页.

其中 $a\geqslant 0$，是某一常数.

［证明］ 由(13)知对任意 x，

$$f(x)=\left[f\left(\frac{x}{2}\right)\right]^2\geqslant 0$$

因此 $f(x)$ 非负. 反复使用(13)，对任意正整数 n 及实数 x 有

$$f(nx)=[f(x)]^n$$

在上式中取 $x=\frac{1}{n}$，得

$$f(1)=\left[f\left(\frac{1}{n}\right)\right]^n$$

记 $a=f(1)\geqslant 0$，则

$$f\left(\frac{1}{n}\right)=a^{\frac{1}{n}}$$

因此，对任意正整数 m 及 n，成立

$$f\left(\frac{m}{n}\right)=\left[f\left(\frac{1}{n}\right)\right]^m=a^{\frac{m}{n}}$$

这样，我们已证得(14)对一切有理数成立，再利用连续性或单调性可以证明对无理数也成立，从而证明了引理.

［普阿松过程］ 考虑来到某交换装置的电话呼叫数，假定它具有下面三个性质：

(i) **平稳性** 在 $[t_0, t_0+t)$ 中来到的呼叫数只与时间间隔长度 t 有关而与时间起点 t_0 无关. 若以 $P_k(t)$ 记在长度为 t 的时间区间中来到 k 个呼叫的概率，当然

$$\sum_{k=0}^{\infty}P_k(t)=1 \tag{15}$$

对任何 $t>0$ 成立.

过程的平稳性表示了它的概率规律不随时间的推移而改变.

(ii) **独立增量性(无后效性)** 在 $[t_0, t_0+t)$ 内来到 k 个呼

则这一事件与时刻 t_0 以前发生的事件独立．换言之，在对时刻 t_0 以前的事件发生情况所作的任何假定之下，计算出来的在$[t_0, t_0+t)$内发生 k 个呼叫的条件概率都等于同一事件的无条件概率．独立增量性表明在互不相交的时间区间内过程进行的相互独立性．

(iii) **普通性** 在充分小的时间间隔中，最多来到一个呼叫．即，若记

$$\psi(t)=\sum_{k=2}^{\infty}P_k(t)=1-P_0(t)-P_1(t) \tag{16}$$

应有 $\psi(t)=o(t)$，即

$$\lim_{t\to 0}\frac{\psi(t)}{t}=0 \tag{17}$$

普通性表明，在同一时间瞬间来两个或两个以上呼叫实际上是不可能的．

下面我们求 $P_k(t)$．

对$\Delta t>0$，考虑$[0, t+\Delta t)$中来到 k 个呼叫的概率 $P_k(t+\Delta t)$，由独立增量性及全概率公式

$$\begin{aligned}P_k(t+\Delta t)=&P_k(t)P_0(\Delta t)+P_{k-1}(t)P_1(\Delta t)+\cdots\\&+P_0(t)P_k(\Delta t),\quad k\geqslant 0\end{aligned} \tag{18}$$

(对 $n\geqslant 1$，假定 $P_{-n}(t)=0$.)

特别地

$$P_0(t+\Delta t)=P_0(t)P_0(\Delta t) \tag{19}$$

$P_0(t)$表示在长度为 t 的时间间隔中没有来呼叫的概率，因此它关于 t 单调下降，由引理知

$$P_0(t)=a^t$$

其中 $a\geqslant 0$，若 $a=0$，则 $P_0(t)\equiv 0$，这说明在不管怎么短的时间间隔内都要来呼叫，因此在有限时间间隔中要来无穷多个呼叫，这种情形不在我们的考虑之列．此外，因 $P_0(t)$是概率，故应有 $a\leqslant 1$，

而当 $a=1$ 时，$P_0(t)\equiv 1$，这表明永不来呼叫，也不是我们感兴趣的情形，所以应有 $0<a<1$，从而存在 $\lambda>0$，使

$$P_0(t)=e^{-\lambda t} \tag{20}$$

因此当 $\Delta t\to 0$ 时，我们有

$$P_0(\Delta t)=e^{-\lambda\Delta t}=1-\lambda\Delta t+o(\Delta t)$$

$$P_1(\Delta t)=1-P_0(\Delta t)-\psi(\Delta t)=\lambda\Delta t+o(\Delta t)$$

$$\sum_{l=2}^{\infty}P_{k-l}(t)P_l(\Delta t)\leqslant\sum_{l=2}^{\infty}P_l(\Delta t)=\psi(\Delta t)=o(\Delta t)$$

故由(18)得

$$P_k(t+\Delta t)=P_k(t)(1-\lambda\Delta t)+P_{k-1}(t)\cdot\lambda\Delta t+o(\Delta t)$$

$$k\geqslant 1$$

因此

$$\frac{P_k(t+\Delta t)-P_k(t)}{\Delta t}=\lambda[P_{k-1}(t)-P_k(t)]+o(1)$$

$$k\geqslant 1$$

令 $\Delta t\to 0$，得

$$P_k'(t)=\lambda[P_{k-1}(t)-P_k(t)],\quad k\geqslant 1 \tag{21}$$

由于已知 $P_0(t)=e^{-\lambda t}$，故有 $P_1'(t)=\lambda[e^{-\lambda t}-P_1(t)]$，可解得 $P_1(t)=\lambda te^{-\lambda t}$，这样下去，可解得一切 $P_k(t)$.

$$P_k(t)=\frac{(\lambda t)^k}{k!}e^{-\lambda t},\quad k=0,1,2,\cdots \tag{22}$$

这正是普阿松分布，参数为 λt.

在随机过程部分，将把这里所得的结果及所用的方法大大推广。

第二章小结

本章中我们讨论了(统计)独立性的概念，它是概率论中最重

要的概念之一．独立性是概率论特有的概念，它的引进大大推动了概率论的发展，前期概率论中最重要的一些结果大都是在独立性的假定下获得的，只有到了近代才开始研究一些不独立但常在另一种较弱独立性假定下的概率模型．读者能在本书的后面部分充分体验到独立性的重要．

我们定义了事件的独立性与试验的独立性，同学们应把它们进行对比．此外，两个事件独立性的定义与多个事件独立性的定义形式上也有区别，应理解为什么会发生这种情况．

条件概率是概率论中另一重要概念，它与独立性有密切联系．在不具有独立性的场合，它将扮演重要角色．条件概率也是某种概率．

本章还导出几个基本公式：乘法定理及其推广形式，全概率公式，贝叶斯公式，它们以后经常被用到．

贝努里试验是概率论中最重要的概型之一．正是通过对这个概型的不断深入地研究，逐渐提出了概率论特有的课题，创造出相应的工具与方法．它对后来的发展有着不可估量的影响．贝努里试验概型在应用上也很重要．在第五章，我们将继续讨论这个概型，证明有关的极限定理并最后解决一些本章遗留的计算问题．

二项分布与普阿松分布是概率论中最重要的三个分布中的两个，另一个分布——正态分布，将在下章出现．这两个分布在理论研究和实际应用中都很重要．二项分布广泛应用于抽样检查，而普阿松分布则大量出现于社会生活和物理现象中．普阿松过程的结构分析是随机过程理论的最基本成果之一．

到目前为止，我们课程的展开基本上还是与概率论的历史发展相平行的，而且，我们已经为概率论另一重要概念——随机变量的引入作了不少的准备．

习　题

1. 把字母 M、A、X、A、M 分别写在一张卡片上，充分混和后重新排列，问正好得到顺序 MAXAM 的概率是多少？

2. 有三个孩子的家庭中，已知有一个是女孩，求至少有一个男孩的概率。

3. 若M件产品中包含m件废品，今在其中任取两件，求：(1)已知取出的两件中有一件是废品的条件下，另一件也是废品的条件概率；(2)已知两件中有一件不是废品的条件下，另一件是废品的条件概率；(3)取出的两件中至少有一件是废品的概率。

4. 袋中有a只黑球，b只白球，甲乙丙三人依次从袋中取出一球(取后不放回)，试分别求出三个人各自取得白球的概率($b\geqslant 3$)。

5. 从$\{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$中随机地取出两个数字，求其和大于 10 的概率。

6. 甲袋中有a只白球，b只黑球，乙袋中有α只白球，β只黑球，某人从甲袋中任取两球投入乙袋，然后在乙袋中任取两球，问最后取出的两球全为白球的概率是多少？

7. 设有N个袋子，每个袋子中装有a只黑球，b只白球，从第一袋中取出一球放入第二袋中，然后从第二袋中取出一球放入第三袋中，如此下去，问从最后一个袋中取出一黑球的概率是多少？

8. 飞机有三个不同的部分遭到射击，在第一部分被击中一弹或第二部分被击中两弹，或第三部分被击中三弹时，飞机才能被击落，其命中率与每一部分的面积成正比，设三个部分的面积的百分比为 0.1, 0.2, 0.7。若已中两弹，求击落飞机的概率。

9. 投硬币n回，第一回出正面的概率为c，第二回后每次出现与前一次相同表面的概率为p，求第n回时出正面的概率，并讨论当$n\to\infty$时的情况。

10. 甲乙两袋各装一只白球一只黑球，从两袋中各取出一球相交换放入另一袋中，这样进行了若干次。以p_n，q_n，r_n分别记在第n次交换后甲袋中将包含两只白球，一只白球一只黑球，两只黑球的概率。试导出p_{n+1}，q_{n+1}，r_{n+1}用p_n，q_n，r_n表出的关系式，利用它们求p_{n+1}，q_{n+1}，r_{n+1}，并讨论当$n\to\infty$时的情况。

11. 设一个家庭中有 n 个小孩的概率为

$$p_n=\begin{cases}\alpha p^n, & n\geqslant 1\\ 1-\dfrac{\alpha p}{1-p}, & n=0\end{cases}$$

这里 $0<p<1$, $0<\alpha<\dfrac{1-p}{p}$, 若认为生一个小孩为男孩或女孩是等可能的, 求证一个家庭有 $k(k\geqslant1)$ 个男孩的概率为 $2\alpha p^k/(2-p)^{k+1}$.

12. 在上题假定下: (1)已知家庭中至少有一个男孩, 求此家庭至少有两个男孩的概率; (2)已知家庭中没有女孩, 求正好有一个男孩的概率.

13. 已知产品中 96% 是合格的, 现有一种简化的检查方法, 它把真正的合格品确认为合格品的概率为 0.98, 而误认废品为合格品的概率为 0.05, 求以简化法检查下为合格品的一个产品确实是合格的概率.

14. 炮战中, 在距目标 250 米, 200 米, 150 米处射击的概率分别为 0.1, 0.7, 0.2, 而在各该处射击时命中目标的概率分别为 0.05, 0.1, 0.2, 现在已知目标被击毁, 求击毁目标的炮弹是由距目标 250 米处射出的概率.

15. 在通讯渠道中, 可传送字符 AAAA, BBBB, CCCC 三者之一, 假定传送这三者的概率分别为 0.3, 0.4, 0.3, 由于通道噪声的干扰, 正确接收到被传送字母的概率为 0.6, 而接收到其他字母的概率是 0.2, 假定前后字母是否被歪曲互不影响, 若接收到的是 ABCA, 问被传送的是 AAAA 的概率.

16. 设 A, B, C 三事件相互独立, 求证 $A\cup B$, AB, $A-B$ 皆与 C 独立.

17. 若 A, B, C 相互独立, 则 $\bar{A}, \bar{B}, \bar{C}$ 亦相互独立.

18. 证明: 事件 $A_1, A_2, \cdots, A_n$ 相互独立的充要条件是下列 2^n 个等式成立:

$$P(\hat{A}_1\hat{A}_2\cdots\hat{A}_n)=P(\hat{A}_1)P(\hat{A}_2)\cdots P(\hat{A}_n)$$

其中 $\hat{A}_i$ 取 A_i 或 $\bar{A}_i$.

19. 若 A 与 B 独立, 证明 $\{\varnothing, A, \bar{A}, \Omega\}$ 中任何一个事件与 $\{\varnothing, B, \bar{B}, \Omega\}$ 中任何一个事件是相互独立的.

20. 对同一目标进行三次独立射击, 第一, 二, 三次射击的命中概率分别为 0.4, 0.5, 0.7, 试求(1)在这三次射击中恰好有一次击中目标的概率; (2)至少有一次命中目标的概率.

21. 设 $A_1, A_2, \cdots, A_n$ 相互独立, 而 $P(A_k)=p_k$, 试求: (1)所有事件全不发生的概率; (2) 诸事件中至少发生其一的概率; (3) 恰好发生其一的概

率.

22. 当元件K或者元件K_1及K_2都发生故障时电路断开，元件K发生故障的概率等于0.3，而元件K_1，K_2发生故障的概率各为0.2，求电路断开的概率.

23. 说明"重复独立试验中，小概率事件必然发生"的确切意思.

24. 在第一台车床上制造一级品零件的概率等于0.7，而在第二台车床上制造此种零件的概率等于0.8，第一台车床制造了两个零件，第二台制造了三个零件，求所有零件均为一级品的概率.

25. 设实验室器皿中产生甲类细菌与乙类细菌的机会是相同的，若某次发现产生了$2n$个细菌，求(1)至少有一个甲类细菌的概率；(2)甲，乙两类细菌各占其半的概率.

26. 掷硬币出现正面的概率为p，掷了n次，求下列概率：(1)至少出现一次正面；(2)至少出现两次正面.

27. 甲，乙，丙三人进行某项比赛，若三个胜每局的概率相等，比赛规定先胜三局者为整场比赛的优胜者，若甲胜了第一、三局，乙胜了第二局，问丙成为整场比赛优胜者的概率是多少？

28. 甲，乙均有n个硬币，全部掷完后分别计算掷出的正面数，试求两人掷出的正面数相等的概率.

29. 在贝努里试验中，事件A出现的概率为p，求在n次独立试验中事件A出现奇数次的概率.

30. 在贝努里试验中，若A出现的概率为p，求在出现m次$\bar{A}$之前出现k次A的概率.

31. 甲袋中有$N-1$只白球和1只黑球，乙袋中有N只白球，每次从甲，乙两袋中分别取出一只球并交换放入另一袋中去，这样经过了n次，问黑球出现在甲袋中的概率是多少，并讨论$n\to\infty$时的情况.

32. 一个工厂出产的产品中废品率为0.005，任意取来1000件，试计算下面概率：(1) 其中至少有两件废品；(2) 其中不超过5件废品；(3) 能以90%的概率希望废品件数不超过多少？

33. 某交往式计算机有20个终端，这些终端被各单位独立操作，使用率各为0.7，求有10个或更多个终端同时操作的概率.

34. 设每次射击打中目标的概率等于0.001，如果射击5000次，试求打中两弹或两弹以上的概率.

35. 某厂长有 7 个顾问，假定每个顾问贡献正确意见的百分比为 0.6，现为某事可行与否而个别征求各顾问意见，并按多数人的意见作出决策，求作出正确决策的概率.

36. 实验室器皿中产生甲，乙两类细菌的机会是相等的，且产生 k 个细菌的概率为

$$p_k=\frac{\lambda^k}{k!}e^{-\lambda},\quad k=0,1,2,\cdots$$

试求：(1) 产生了甲类细菌但没有乙类细菌的概率；(2) 在已知产生了细菌而且没有甲类细菌的条件下，有 2 个乙类细菌的概率.

37. 假定人在一年 365 日中的任一日出生的概率是一样的，在 50 个人的单位中有两个以上的人生于元旦的概率是多少？

38. 一本 500 页的书，共有 500 个错字，每个字等可能地出现在每一页上，试求在给定的一页上至少有三个错字的概率.

39. 某商店中出售某种商品，据历史记录分析，每月销售量服从普阿松分布，参数为 7，问在月初进货时要库存多少件此种商品，才能以 0.999 的概率充分满足顾客的需要.

40. 螺丝钉生产中废品率为 0.015，问一盒应装多少只才能保证每盒中有 100 只以上的好螺丝钉的概率不小于 80%（提示：用普阿松逼近，设应装 $100+k$ 只）.

41. 某疫苗中所含细菌数服从普阿松分布，每 1 毫升中平均含有一个细菌，把这种疫苗放入 5 只试管中，每试管放 2 毫升，试求：(1)5 只试管中都有细菌的概率；(2)至少有 3 只试管中有细菌的概率.

42. 通过某交叉路口的汽车流可看作普阿松过程，若在一分钟内没有车的概率为 0.2，求在 2 分钟内有多于一车的概率.

43. 若每蚕产 n 个卵的概率服从普阿松分布，参数为 λ，而每个卵变为成虫的概率为 p，且各卵是否变为成虫彼此间没有关系，求每蚕养出 k 只小蚕的概率.

44. 若已知 $t=0$ 时，某分子与另一分子碰撞，又知对任何 $t\geqslant 0$ 和 $\Delta t>0$，若不管该分子在时刻 t 以前是否遭受碰撞，在 $(t,t+\Delta t)$ 中遭到碰撞的概率等于 $\lambda\Delta t+o(\Delta t)$，试求该分子在时刻 τ 还没有再受到碰撞的概率.

*45. 利用概率论的想法证明下面恒等式：

$$\sum_{k=0}^{N}\binom{N+k}{k}\frac{1}{2^k}=2^N$$

*46. 通过构造适当的概率模型证明：从正整数中随机地选取两数，此两数互素的概率等于$\frac{6}{\pi^2}$.

47. 某车间宣称自己产品的合格率超过99%，检验人员从该车间的10000件产品中抽查了100件，发现有两件次品，能否据此断定该车间谎报合格率？

第三章　随机变量与分布函数

§1. 随机变量及其分布

一、随机变量的定义

在随机现象中，有很大一部分问题与数值发生关系，例如在产品检验问题中，我们关心的是抽样中出现的废品数；在车间供电问题中我们关心的是某时刻正在工作的车床数；在电话问题中关心的是某段时间中的话务量，它与呼叫的次数及各次呼叫占用交换设备的时间长短有关. 此外如测量时的误差，气体分子运动的速度，信号接收机所收到的信号(用电压表示或数字表示)的大小，也都与数值有关.

有些初看起来与数值无关的随机现象，也常常能联系数值来描述，例如在掷硬币问题中，每次出现的结果为正面或反面，与数值没有关系，但是我们能用下面方法使它与数值联系起来，当出现正面时对应数“1”，而出现反面时对应数“0”，为了计算 n 次投掷中出现的正面数就只须计算其中“1”出现的次数了.

一般地，如果 A 为某个随机事件，则一定可以通过如下示性函数使它与数值发生联系：

$$I_A=\begin{cases}1, & \text{如果} A \text{发生}\\ 0, & \text{如果} A \text{不发生}\end{cases}$$

这些例子中，试验的结果能用一个数 ξ 来表示，这个数 ξ 是随着试验的结果的不同而变化的，也即它是样本点的一个函数，这种量以后称为**随机变量**. 本书中将用希腊字母 $\xi, \eta, \zeta, \cdots$ 等来表示随机变量. 下面我们就来考虑应当如何给这种量以严格的数学

定义.

正如对随机事件一样，我们所关心的不仅是试验会出现什么结果，更重要的是要知道这些结果将以怎样的概率出现，也即对随机变量我们不但要知道它取什么数值，而且要知道它取这些数值的概率.

从随机现象的可能出现的结果来看，随机变量至少有两种不同的类型. 一种是试验结果 ξ 所可能取的值为有限个或至多可列个，我们能把其可能结果一一列举出来，这种类型的随机变量称为**离散型随机变量**. 在日常生活中经常碰到离散型随机变量，例如废品数，电话呼叫数等等. 前面讨论过的随机现象大部分都能用离散型随机变量来描述. 例如古典概型中只有有限个可能结果，若对应于每一个结果用一个数值来表征，则得到一个离散型随机变量. 又如在 n 次贝努里试验中，若以 μ 记事件 A 出现的次数，则 μ 可取值 $0, 1, 2, \cdots, n$. 在呼叫流的研究中，若以 $\xi(t)$ 记 $[0, t]$ 中来到的呼叫数，则 $\xi(t)$ 可取值 $0, 1, 2, \cdots$，这些都是离散型随机变量.

从上章的讨论可以看到，要描述这类随机变量并不难，以 n 次贝努里试验中事件 A 出现的次数 μ 为例，我们知道它是样本点 ω 的函数，也就是说严格来讲，应写作 $\mu(\omega)$，其中 $\omega\in\Omega$，它取的值是 $0, 1, 2, \cdots, n$，并且知道 $\mu(\omega)$ 取这些值的概率为①

$$P\{\mu(\omega)=k\}=\binom{n}{k}p^k q^{n-k},\quad k=0, 1, 2, \cdots, n$$

这里我们已经知道了 $\mu(\omega)$ 取什么值，以及以什么概率取这些值.

---

① 下面出现记号 $\{\mu(\omega)=k\}$，对于不熟悉这类记号的读者，我们作如下说明. $\{\mu(\omega)=k\}$ 是 $\{\omega:\ \mu(\omega)=k\}$ 的简写，它表示具有如下性质的样本点 ω 的集合：在其上 μ 取固定值 k. 一般地，对于从样本空间 Ω 到数直线 R^1 上的单值映照 $\xi(\omega)$，若 A 是 R^1 的某一子集，常以 $\{\xi(\omega)\in A\}$ 作为 $\{\omega:\ \xi(\omega)\in A\}$ 的简写，用来表示使 $\xi(\omega)$ 之值属于 A 的那些样本点 ω 的集合，有时还进一步简记作 $\xi^{-1}(A)$；特别地，$\{\xi(\omega)<x\}$ 即为 $\{\omega:\ \xi(\omega)\in(-\infty, x)\}$，这里 x 是某一实数.

一般地，对于定义在样本空间Ω上的离散型随机变量$\xi(\omega)$只要能指出它取的值$x_1, x_2\cdots, x_n, \cdots$，以及它取这些值的概率$P\{\xi(\omega)=x_i\}, i=1,2,\cdots,n,\cdots$，就满足了我们的要求．显然要做到这一点，必须要求$\{\xi(\omega)=x_i\}$有概率．因为我们只对事件域$\mathscr{F}$中的集合定义概率，所以必须有$\{\omega:\xi(\omega)=x_i\}\in\mathscr{F}$．

与离散型随机变量不同，一些随机现象所出现的试验结果ξ不止取可列个值，例如测量误差，分子运动速度，候车时的等待时间，降水量，风速，洪峰值等等皆是．这时用来描述试验结果的随机变量还是样本点ω的函数：严格写应是$\xi(\omega)$，其中$\omega\in\Omega$，但是这随机变量能取某个区间$[c,d]$或$(-\infty,\infty)$的一切值．

假如想用描述离散型随机变量的方法（简单地罗列所取的值及相应的概率）来描述这后一类随机变量，则会碰到很大的困难．一来是这类随机变量所取的值不能一一列出；二来，我们下面将会看到，取连续值的随机变量，它取某个特定值的概率常常是0，因此用这种描述方法根本不行．

对于取连续值的随机变量我们所关心的也并不是它取某个特定值的概率，例如在测量误差的讨论中，我们感兴趣的是测量误差小于某个数的概率；在降雨问题中，我们重视的是雨量在某一个量级，例如在100毫米到120毫米之间的概率．总之，对于取连续值的随机变量$\xi(\omega)$，我们感兴趣的是$\xi(\omega)$取值于某个区间(a,b)的概率，或取值于若干个这种区间的概率．因此应当要求$\{a\leqslant\xi(\omega)<b\}$或$\{\xi(\omega)<b\}$或一般地$\{\xi(\omega)\in A\}$（其中$A$是由区间经并，交运算而得到的直线上的某一个点集）有概率可言，既然只对概率空间$(\Omega,\mathscr{F},P)$的事件域$\mathscr{F}$中的集合才定义概率，因此我们自然要求上述集合属于$\mathscr{F}$，即都是事件．

通过上面讨论可以看出，为了使我们感兴趣的概率计算得以进行，我们应对$\xi(\omega)$加上一定的限制，主要是要求$\{\omega:\xi(\omega)\in B\}$

应是事件．在离散型随机变量的场合，B是直线上的某一个点；在取连续值场合，B是直线上由区间经并、交运算而得到的某一点集．在概率计算中有时要考虑可列运算，因此较方便的是取B为直线上波雷尔点集．

为此引进如下定义：

定义 设$\xi(\omega)$是定义于概率空间$(\Omega, \mathscr{F}, P)$上的单值实函数，如果对于直线上任一波雷尔点集B，有

$$\{\omega:\xi(\omega)\in B\}\in\mathscr{F} \tag{1}$$

则称$\xi(\omega)$为**随机变量**（图1）．而$P\{\xi(\omega)\in B\}$称为随机变量$\xi(\omega)$的**概率分布**．

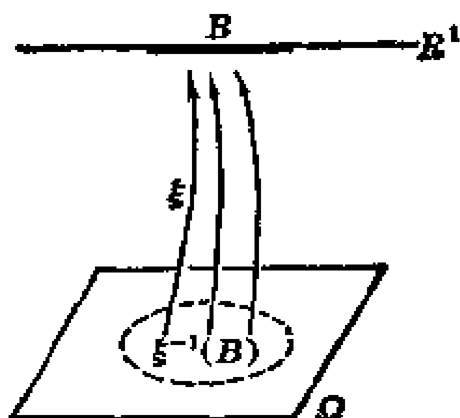


图1 随机变量

特别地，若取$B=(-\infty, x)$，则有

$$\{\omega:\xi(\omega)<x\}\in\mathscr{F} \tag{2}$$

因此$P\{\xi(\omega)<x\}$有定义．注意到

$$P\{a\leqslant\xi(\omega)<b\}=P\{\xi(\omega)<b\}-P\{\xi(\omega)<a\} \tag{3}$$

所以只要对一切实数x给出概率$P\{\xi(\omega)<x\}$，就能算出$\xi(\omega)$落入某个区间$[a, b)$的概率，再利用概率的性质还可以算出$\xi(\omega)$属于某些相当复杂的直线点集（譬如可列个不相交的左闭右开区间之和）的概率．

定义 称

$$F(x)=P\{\xi(\omega)<x\}, \quad -\infty<x<\infty \tag{4}$$

为随机变量$\xi(\omega)$的**分布函数**．

由(3)立刻得到

$$P\{a\leqslant\xi(\omega)<b\}=F(b)-F(a) \tag{5}$$

在上面讨论中，我们根据描述随机变量的需要而给出了随机变量与分布函数的定义．应该指出，最后所以采用这种定义，还有数学理论上的需要，有些进一步的事实，例如若(2)成立则(1)也成

立，由分布函数可以唯一决定概率分布等等是可以通过测度论的方法证明的，这些已超出本课程范围，我们将不予讨论.

概率论是研究大量随机现象中的数量规律的数学分支，研究随机变量和分布函数是它的重要任务，而且概率论中所研究的也大都局限于能用随机变量来描述的随机现象. 前面已指出，随机事件的研究可通过示性函数转化为对随机变量的研究，随机变量是取数值的，因此可以对它进行各种数学运算，研究起来就很方便.

二、分布函数的性质

我们先把分布函数的最基本的性质汇集于下列定理中.

定理 分布函数 $F(x)$ 具有下列性质:

(i) 单调性：若 $a<b$，则 $F(a)\leqslant F(b)$;

(ii) $\lim\limits_{x\to-\infty}F(x)=0,\ \lim\limits_{x\to+\infty}F(x)=1;$ **(6)**

(iii) 左连续性：$F(x-0)=F(x)$.

[证明](i) $F(b)-F(a)=P\{a\leqslant\xi<b\}\geqslant 0$.

$$
\begin{aligned}
\text{(ii)}\quad P\{-\infty<\xi<\infty\}&=\sum_{n=-\infty}^{\infty}P\{n\leqslant\xi<n+1\}\\
&=\sum_{n=-\infty}^{\infty}[F(n+1)-F(n)]\\
&=\lim_{n\to+\infty}F(n)-\lim_{m\to-\infty}F(m)=1
\end{aligned}
$$

由于 $F(x)$ 的单调性：$\lim\limits_{x\to-\infty}F(x)=\lim\limits_{m\to-\infty}F(m)$，$\lim\limits_{x\to+\infty}F(x)=\lim\limits_{n\to+\infty}F(n)$ 存在.

因为 $0\leqslant F(x)\leqslant 1$，故

$$\lim_{x\to-\infty}F(x)=0,\quad \lim_{x\to+\infty}F(x)=1$$

今后为书写方便起见，将记

$$F(-\infty)=\lim_{x\to-\infty}F(x),\quad F(+\infty)=\lim_{x\to+\infty}F(x) \tag{7}$$

(iii) 由于 $F(x)$是单调函数，只须证明对于一列单调上升的数列 $x_0<x_1<x_2<\cdots<x_n<\cdots, x_n\to x$ 成立$\lim\limits_{n\to\infty}F(x_n)=F(x)$即可.

因为

$$F(x)-F(x_0)=P\{x_0\leqslant\xi<x\}=\sum_{n=1}^{\infty}[F(x_n)-F(x_{n-1})]$$

$$=\lim_{n\to\infty}F(x_n)-F(x_0)$$

所以

$$F(x-0)=\lim_{n\to\infty}F(x_n)=F(x)$$

可以看出分布函数的这三个基本性质，正好对应于概率的三个基本性质.

有了分布函数，关于随机变量 $\xi(\omega)$ 的许多概率都能方便算出，例如

$$P\{\xi(\omega)=a\}=F(a+0)-F(a) \tag{8}$$

$$P\{\xi(\omega)\leqslant a\}=F(a+0) \tag{9}$$

$$P\{\xi(\omega)\geqslant a\}=1-F(a) \tag{10}$$

$$P\{\xi(\omega)>a\}=1-F(a+0) \tag{11}$$

综上所述，分布函数是一种分析性质良好的函数，便于处理，而给定了分布函数就能算出各种事件的概率，因此引进分布函数使许多概率论问题得以简化而归结为函数的运算，这样就能利用数学分析的许多结果，这也是引进随机变量的好处之一.

可以证明满足上述定理中三个性质的函数必是某随机变量的分布函数，关于这点的讨论将在第 3 节进行.

对于随机变量及其概率分布的研究，若按其分布函数的不同类型分别讨论是有好处的，下面我们将照此线索进行.

三、离散型随机变量

设$\{x_i\}$为离散型随机变量ξ的所有可能值；而$p(x_i)$是ξ取x_i的概率，即

$$P\{\xi=x_i\}=p(x_i),\quad i=1,2,3,\cdots \tag{12}$$

$\{p(x_i),\ i=1,2,3,\cdots\}$称为随机变量$\xi$的**概率分布**，它应满足下面关系：

$$p(x_i)\geqslant 0,\quad i=1,2,3,\cdots \tag{13}$$

$$\sum_{i=1}^{\infty}p(x_i)=1 \tag{14}$$

当给定了$\{x_i,\ i=1,2,\cdots\}$及$\{p(x_i),\ i=1,2,\cdots\}$，我们就能很好地描述随机变量$\xi$，因为我们已经知道了它取什么值，以及以什么概率取这些值，而这正是我们对随机变量所关心的问题。

常用下面方法来表出离散型随机变量ξ的概率分布，

$$\begin{pmatrix} x_1, & x_2, & \cdots, & x_n, & \cdots \\ p(x_1), & p(x_2), & \cdots, & p(x_n), & \cdots \end{pmatrix} \tag{15}$$

(15)称为随机变量ξ的**分布列**，由分布列能一目了然地看出随机变量ξ的取值范围及取这些值的概率。

有了分布列，可以通过下式求得分布函数

$$F(x)=P\{\xi<x\}=\sum_{x_k<x}p(x_k) \tag{16}$$

显然这时$F(x)$是一个跳跃函数，它在每个x_k处有跳跃度$p(x_k)$。当然，由$F(x)$也可唯一决定x_k及$p(x_k)$，因此用分布列或分布函数都能描述离散型随机变量。

下面看一些离散型随机变量及其概率分布的例子。这些分布大都在上两章引入过。

［退化分布］若随机变量α只取常数值c，即

$$P\{\alpha=c\}=1$$

这时分布函数为

$$I(x-c)=\begin{cases}0, & x\leqslant c\\ 1, & x>c\end{cases} \tag{17}$$

可以说，这样的 α 并不随机，但有时我们宁愿把它看作随机变量的退化情况更为方便，因此称之为**退化分布**，又称**单点分布**.

[贝努里分布] 在一次试验中，事件 A 出现的概率为 p，不出现的概率为 $q=1-p$，若以 β 记事件 A 出现的次数，则 β 仅取 0，1 二值，相应的分布列为

$$\begin{pmatrix}0, & 1\\ q, & p\end{pmatrix} \tag{18}$$

这个分布称为**贝努里分布**亦称**两点分布**.

[二项分布] 在 n 重贝努里试验中，若以 μ 记事件 A 出现的次数，则它是一个随机变量，μ 可能取的值为 $0, 1, 2, \cdots, n$，其对应的概率由**二项分布**给出

$$b(k; n, p)=P\{\mu=k\}=\binom{n}{k}p^k q^{n-k}$$

$$k=0, 1, 2, \cdots, n \tag{19}$$

关于二项分布及其计算已在上章作过详细讨论，这里不再重复，至于为什么称为分布，到了此处自然十分明白.

顺便指出，贝努里分布可以看作 $n=1$ 时的二项分布，这时相应于一次独立试验的场合.

[超几何分布] 对某批 N 件产品进行不放回抽样检查，若这批产品中有 M 件次品，现从整批产品中随机抽出 n 件产品，则在这 n 件产品中出现的次品数 ν 是随机变量，它取值 $0, 1, 2, \cdots, n$，其概率分布为**超几何分布**.

$$h_k=P\{\nu=k\}=\frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}, \quad \begin{matrix}0\leqslant k\leqslant n\leqslant N\\ k\leqslant M\end{matrix} \tag{20}$$

这个分布在第一章中已经出现，我们还证明过当N很大而n较小时，它可用二项分布来近似.

［普阿松分布］ 若随机变量ξ可取一切非负整数值，且

$$P\{\xi=k\}=\frac{\lambda^k}{k!}e^{-\lambda},\quad k=0,1,2,\cdots \tag{21}$$

其中$\lambda>0$，则称ξ服从**普阿松分布**.

普阿松分布是概率论中非常重要的一个分布，在第二章§4中已对它进行过初步讨论，以后我们还会经常提到它.

［几何分布］ 在事件A发生的概率为p的贝努里试验中，若以η记A首次出现时的试验次数，则η为随机变量，它可能取的值为1,2,3…，其概率分布为**几何分布**.

$$g(k,p)=P\{\eta=k\}=q^{k-1}p,\quad k=1,2,\cdots \tag{22}$$

作为一种等待分布，我们在前两章中已多次碰到过几何分布，这也说明了几何分布是一种常见的概率分布；几何分布在概率论中的重要性，还在于它具有下面特殊的性质.

几何分布的无记忆性 如上所述，在贝努里试验中，等待首次成功的时间η服从几何分布(22). 现在假定已知在前m次试验中没有出现成功，那么为了达到首次成功所再需要的等待时间η'也还是服从几何分布(22)，与前面的失败次数m无关，形象化地说，就是把过去的经历完全忘记了. 因此无记忆性是几何分布所具有的一个有趣的性质. 但是更加有趣的是，在离散型分布中，也只有几何分布才具有这样一种特殊的性质. 下面我们来严格叙述并证明这个事实.

若η是取正整数值的随机变量，并且，在已知$\eta>k$的条件下，$\eta=k+1$的概率与k无关，那么η服从几何分布.

这个论断的证明如下：以p记上述条件概率，并记$q_k=P\{\eta>k\}$及$p_k=P\{\eta=k\}$，那么$p_{k+1}=q_k-q_{k+1}$，而且在已知$\eta>k$

的条件下，$\eta=k+1$ 的条件概率为$\frac{p_{k+1}}{q_k}$，因此

$$\frac{p_{k+1}}{q_k}=p$$

即

$$\frac{q_{k+1}}{q_k}=1-p$$

注意到 $q_0=1$，那么 $q_k=(1-p)^k$，因此

$$p_k=(1-p)^{k-1}p,\quad k=1,2,\cdots$$

这正是几何分布(22).

[巴斯卡分布] 在上例的贝努里试验中，若以 ξ 记第 r 次成功出现时的试验次数，则 ξ 是随机变量，取值 $r, r+1, \cdots$，其概率分布为巴斯卡分布

$$P\{\xi=k\}=\binom{k-1}{r-1}p^r q^{k-r},\quad k=r, r+1, \cdots \tag{23}$$

这分布在上章 § 3出现过. 显然当 $r=1$ 时，即为几何分布. 另一方面若以 η_i 记等待第 i 次成功的试验次数（从第 $i-1$ 次成功之后第一次试验算起至第 i 次成功止），则 η_i 服从几何分布(22)，而且

$$\xi=\eta_1+\cdots+\eta_r \tag{24}$$

四、连续型随机变量

除了离散型随机变量之外，还有一类重要的随机变量——**连续型随机变量**. 这种随机变量 ξ 可取某个区间$[c, d]$或$(-\infty, \infty)$中的一切值，而且其分布函数 $F(x)$是绝对连续函数，即存在可积函数 $p(x)$，使

$$F(x)=\int_{-\infty}^{x}p(y)dy \tag{25}$$

称 $p(x)$为 ξ 的(分布)**密度函数**.

由分布函数的性质可知对 $p(x)$应有

$$p(x)\geqslant 0 \tag{26}$$

$$\int_{-\infty}^{\infty} p(x)dx=1 \tag{27}$$

反之，对于定义在$(-\infty,\infty)$上的可积函数 $p(x)$，若它满足(26)及(27)，则由(25)定义的函数 $F(x)$是一个分布函数，即它有分布函数所必须具备的三个性质.

由(5)立刻得到

$$P\{a\leqslant\xi<b\}=F(b)-F(a)=\int_a^b p(x)dx \tag{28}$$

因此给定密度函数，便可以算出随机变量落入某一个区间的概率.

下面计算 $P\{\xi=c\}$，因为

$$P\{\xi=c\}\leqslant P\{c\leqslant\xi<c+h\}=\int_c^{c+h} p(x)dx$$

故

$$0\leqslant P\{\xi=c\}\leqslant\lim_{h\to 0}\int_c^{c+h} p(x)dx=0$$

因此

$$P\{\xi=c\}=0 \tag{29}$$

即连续型随机变量取个别值的概率为 0，这与离散型随机变量截然不同. 因此用列举连续型随机变量取某个值的概率来描述这种随机变量不但做不到，而且也毫无意义.

此外，上述结果还表明，一个事件的概率等于零，这事件并不一定是不可能事件；同样地，一个事件的概率等于 1，这事件也不一定是必然事件.

由于

$$p(x)\Delta x\approx\int_x^{x+\Delta x} p(y)dy=F(x+\Delta x)-F(x) \tag{30}$$

因此密度函数 $p(x)$ 的数值反映了随机变量取 x 的邻近的值的概率的大小. 所以用密度函数描述连续型随机变量的概率分布在某

种意义上与离散型时用分布列描述，又有相似之处.

下面举一些常见的连续型分布的例子.

[均匀分布]　若 a, b 为有限数，由下列密度函数定义的分布称为$[a,b]$上**均匀分布**.

$$p(x)=\begin{cases}\dfrac{1}{b-a}, & a\leqslant x\leqslant b\\ 0\ , & x<a\text{ 或 }x>b\end{cases}\tag{31}$$

相应的分布函数为

$$F(x)=\begin{cases}0\ , & x\leqslant a\\ \dfrac{x-a}{b-a}, & a<x\leqslant b\\ 1\ , & x>b\end{cases}\tag{32}$$

$[a,b]$上均匀分布有时简记作 $U[a,b]$.

若随机变量 ξ 服从$[a,b]$上均匀分布，则 ξ 在$[a,b]$中取值落在某一区域内的概率与这个区域的测度成正比. 粗略地讲就是，ξ 取$[a,b]$中任一点的可能性一样. 当然也可以反过来看，均匀分布正是把这种直观的讲法严格化.

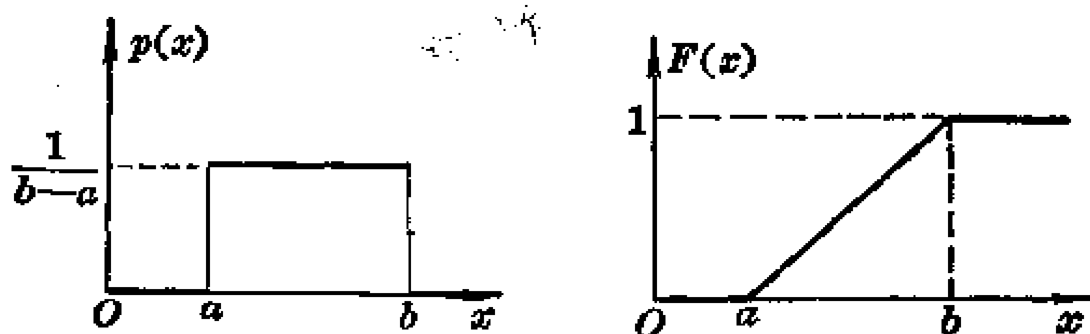


图 2　均匀分布的密度函数与分布函数

图 2 画出了$[a,b]$上均匀分布的密度函数及分布函数.

[例 1]　定点计算中的舍入误差可以作为常见的均匀分布随机变量的例子. 假如我们在运算中，数据都只保留到小数点后第五位，而小数点第五位以后的数字按四舍五入处理. 若以 x 表示

真值，以 $\hat{x}$ 表示舍入后的值，则误差 $e=x-\hat{x}$ 一般假定为$[-0.5\times10^{-5}, 0.5\times10^{-5})$ 上均匀分布的随机变量，有了这个假定，就能对经过大量运算后的数据进行误差分析，这种误差分析在用数字计算机解题时是很必要的，因为数字计算机字长总是有限的。

［正态分布］ 密度函数为

$$p(x)=\frac{1}{\sqrt{2\pi}\,\sigma}e^{-\frac{(x-a)^2}{2\sigma^2}}, \quad -\infty<x<\infty \tag{33}$$

其中 $\sigma>0$，a 与 σ 均为常数，相应的分布函数为

$$F(x)=\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{x}e^{-\frac{(y-a)^2}{2\sigma^2}}dy, \quad -\infty<x<\infty \tag{34}$$

这分布称为正态分布，简记为 $N(a, \sigma^2)$。

特别当 $a=0$，$\sigma=1$，这时分布称为**标准正态分布**，记为 $N(0, 1)$，相应的分布密度函数及分布函数分别记为 $\varphi(x)$ 及 $\Phi(x)$（图 3）。

$$\varphi(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}, \quad -\infty<x<\infty \tag{35}$$

$$\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-\frac{y^2}{2}}dy, \quad -\infty<x<\infty \tag{36}$$

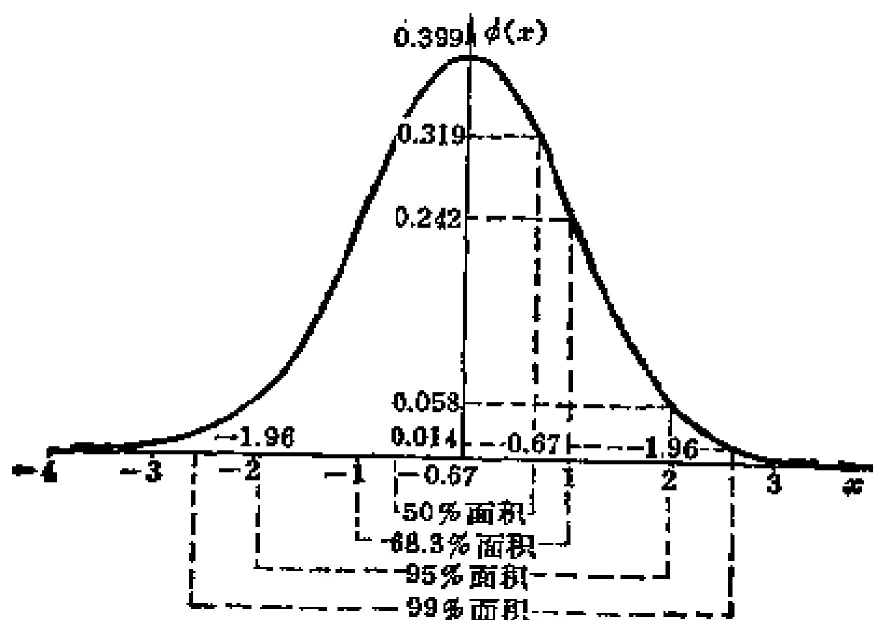


图 3 正态密度函数 $\phi(x)$

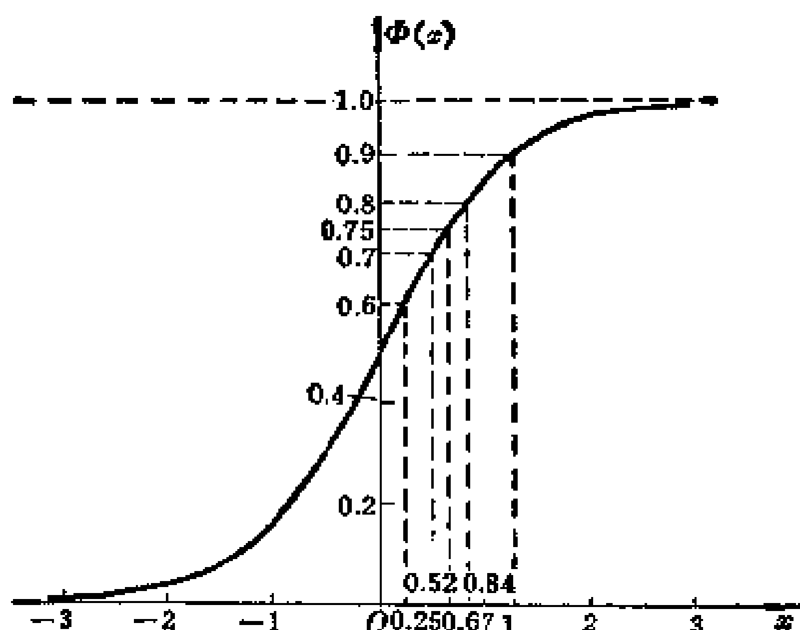


图 3　正态分布函数 $\Phi(x)$

习惯上把服从正态分布的随机变量称为正态变量．可以验证，若随机变量 ξ 服从正态分布 $N(a,\ \sigma^2)$，则随机变量 $\eta=\frac{\xi-a}{\sigma}$ 服从 $N(0,1)$.

为了说明 (33) 确实定义了一个密度函数，需要验证关系式 (26)及(27)，显然

$$p(x)>0$$

因此剩下的是验证(27)．令 $\frac{x-a}{\sigma}=y$，则

$$\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-a)^2}{2\sigma^2}}dx=\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{y^2}{2}}dy$$

但是

$$\left(\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx\right)\left(\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{y^2}{2}}dy\right)$$

$$=\frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}e^{-\frac{x^2+y^2}{2}}dxdy$$

变换到极坐标，令 $x=r\cos\theta,\ y=r\sin\theta$，则

$$\frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}e^{-\frac{x^2+y^2}{2}}dxdy$$

$$=\frac{1}{2\pi}\int_{0}^{\infty}\int_{0}^{2\pi}e^{-\frac{r^2}{2}}rdrd\theta$$

$$=\int_{0}^{\infty}re^{-\frac{r^2}{2}}dr=1$$

由$\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx$的非负性知

$$\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}\,\sigma}e^{-\frac{(x-a)^2}{2\sigma^2}}dx=\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{y^2}{2}}dy=1$$

从而完成了(27)的验证.

正态分布是概率论中最重要的一种分布．一方面，正态分布是自然界最常见的一种分布，例如测量的误差；炮弹弹落点的分布；人的生理特征的尺寸：身长，体重等等；农作物的收获量；工厂产品的尺寸：直径，长度，宽度，高度，……，等等都近似服从正态分布．一般说来，若影响某一数量指标的随机因素很多，而每个因素所起的作用不太大，则这个指标服从正态分布，这点可以利用概率论的极限定理来加以证明．另一方面，正态分布具有许多良好的性质，许多分布可用正态分布来近似，另外一些分布又可以通过正态分布来导出，因此在理论研究中，正态分布十分重要．

一般正态分布密度函数 $p(x)$的图形见图 4，通过图形或(33)不难看出：$p(x)$在 $x=a$ 处达到极大，整个图形关于 $x=a$ 对称．当 σ 不同时，$p(x)$ 的形状也不同，σ 越小，分布越集中在 $x=a$ 附近；当 σ 越大时，分布就越平坦．

由于正态分布在概率计算中的重要性，已编造了各种各样的正态分布表，本书的附表中就有 $a=0$，$\sigma=1$ 时的正态分布密度函数 $\varphi(x)$ 和正态分布函数 $\Phi(x)$ 的表．由于 $\varphi(-x)=\varphi(x)$ 及 $\Phi(-x)=1-\Phi(x)$，所以表中只对正的 x 给出 $\varphi(x)$及$\Phi(x)$的数

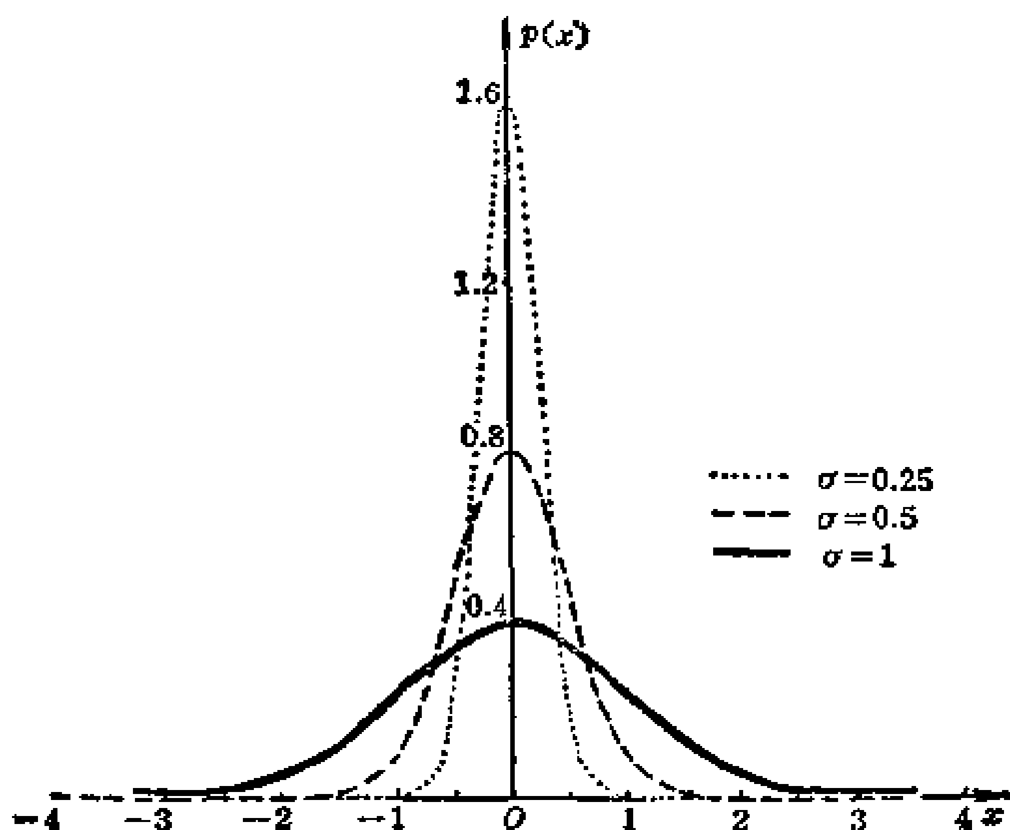


图 4　对于同一个 a 和不同的 σ^2 的正态密度曲线

值. 一般 $N(a, \sigma^2)$ 的分布函数值可由变换而得

$$F(x)=\Phi\left(\frac{x-a}{\sigma}\right) \tag{37}$$

从表中可以看出, 若 ξ 服从 $N(a, \sigma^2)$, 则

$$P\{|\xi-a|<\sigma\}\approx 68.27\% \tag{38}$$

$$P\{|\xi-a|<2\sigma\}\approx 95.45\% \tag{39}$$

$$P\{|\xi-a|<3\sigma\}\approx 99.73\% \tag{40}$$

因此可以说, 在一次试验里, ξ 几乎总是落在 $(a-3\sigma, a+3\sigma)$ 之中.

下例说明正态分布与实测数据符合得很好.

[例 2]　某手表厂曾对其生产的某个零件的重量收集了大量资料, 对测量得的 3805 个数据, 按不同重量加以分组, 并记录了不同范围内零件的个数(频数), 计算了它们的频率, 结果如下表所

示，它们与$a=56.94$，$\sigma=8.2$的正态分布符合得相当好，图5表示了两者符合的情况.

区间 $[x_i, x_{i+1})$	频数 n_i	频率 $f_i=\frac{n_i}{n}$	$\Phi\left(\frac{x_{i+1}-a}{\sigma}\right)-\Phi\left(\frac{x_i-a}{\sigma}\right)$
$[-\infty, 41.5)$	125	0.03285	0.03005
$[41.5, 43.5)$	72	0.01892	0.02150
$[43.5, 45.5)$	124	0.03259	0.02921
$[45.5, 47.5)$	145	0.03811	0.04431
$[47.5, 49.5)$	193	0.05072	0.05630
$[49.5, 50.5)$	137	0.0360	0.0334
$[50.5, 51.5)$	131	0.0344	0.0398
$[51.5, 52.5)$	154	0.0405	0.0400
$[52.5, 53.5)$	156	0.0410	0.0426
$[53.5, 54.5)$	174	0.0457	0.0457
$[54.5, 55.5)$	186	0.0489	0.0472
$[55.5, 56.5)$	191	0.0502	0.0484
$[56.5, 57.5)$	206	0.0541	0.0486
$[57.5, 58.5)$	193	0.0507	0.0482
$[58.5, 59.5)$	185	0.0486	0.0472
$[59.5, 60.5)$	153	0.0402	0.0454
$[60.5, 61.5)$	176	0.0463	0.0430
$[61.5, 62.5)$	147	0.0386	0.0402
$[62.5, 63.5)$	144	0.0378	0.0370
$[63.5, 64.5)$	140	0.0368	0.0331
$[64.5, 65.5)$	109	0.0286	0.0299
$[65.5, 66.5)$	111	0.0292	0.0282
$[66.5, 67.5)$	93	0.02444	0.02247
$[67.5, 69.5)$	127	0.03338	0.03552
$[69.5, 71.5)$	81	0.02129	0.02547
$[71.5, \infty)$	152	0.03995	0.03754
总　　和	$n=3805$		

我们通过下面例子来说明正态分布的计算.

[例 3]　从南郊某地乘车前往北区火车站搭火车有两条路线

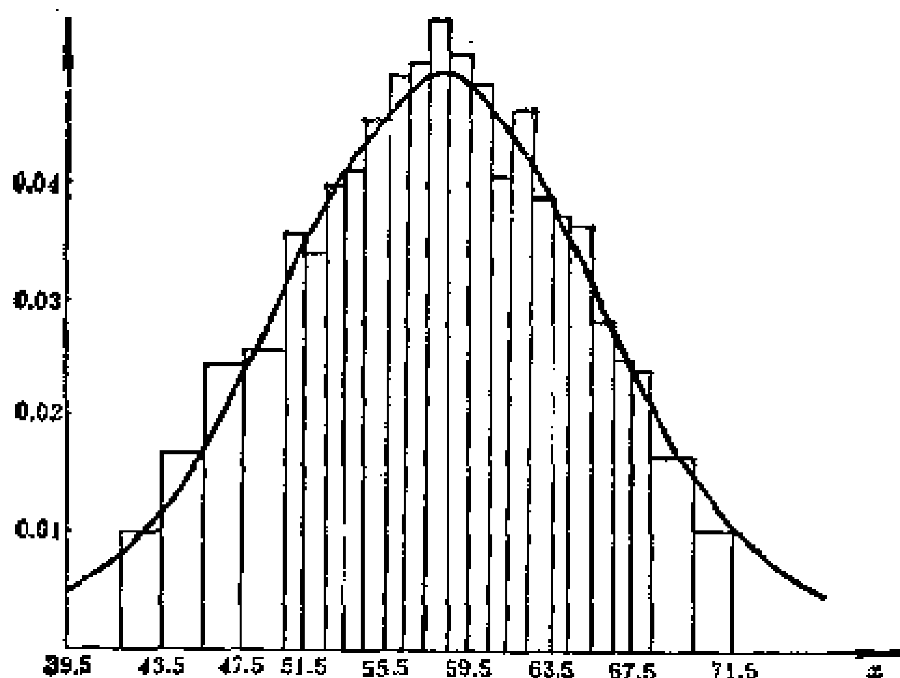


图 5 实测数据频率直方图与正态密度曲线

可走，第一条路线穿过市区，路程较短，但交通拥挤，所需时间(单位为分)服从正态分布 $N(50, 100)$，第二条路线沿环城公路走，路程较长，但意外阻塞较少，所需时间服从正态分布 $N(60, 16)$，(1)假如有 70 分钟可用，问应走哪一条路线？(2)若只有 65 分钟可用，又应走哪一条路线？

[解] 显然应走在允许的时间内有较大概率及时赶到火车站的路线，若以 τ 记行车时间，则有关概率如下：

(1) 有 70 分钟可用时走第一条路线及时赶到的概率为

$$P\{\tau \leqslant 70\} = \Phi\left(\frac{70-50}{10}\right) = \Phi(2) = 0.9772$$

走第二条路线及时赶到的概率为

$$P\{\tau \leqslant 70\} = \Phi\left(\frac{70-60}{4}\right) = \Phi(2.5) = 0.9938$$

因此在这种场合，应走第二条路线.

(2) 只有 65 分钟可用时走第一条路线及时赶到的概率为

$$P\{\tau \leqslant 65\} = \Phi\left(\frac{65-50}{10}\right) = \Phi(1.5) = 0.9332$$

走第二条路线及时赶到的概率为

$$P\{\tau \leqslant 65\} = \Phi\left(\frac{65-60}{4}\right) = \Phi(1.25) = 0.8944$$

因此这种场合应走第一条路线.

[指数分布] 分布密度函数为

$$p(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geqslant 0 \\ 0, & x < 0 \end{cases} \tag{41}$$

分布函数为

$$F(x) = \begin{cases} 1 - e^{-\lambda x}, & x \geqslant 0 \\ 0, & x < 0 \end{cases} \tag{42}$$

这里 $\lambda > 0$, 是常数, 这分布称为**指数分布**.

指数分布有重要应用, 常用它来作为各种"寿命"分布的近似, 例如无线电元件的寿命, 动物的寿命, 电话问题中的通话时间, 随机服务系统中的服务时间等都常假定服从指数分布 .

指数分布的重要性还表现在它具有类似于几何分布的"无记忆性". 设随机变量 ξ 服从指数分布(42), 则对于任意的 $s > 0$, $t > 0$,

$$P\{\xi > s + t \mid \xi > s\} = \frac{P\{\xi > s + t\}}{P\{\xi > s\}} = \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} = e^{-\lambda t}$$

因此

$$P\{\xi > s + t \mid \xi > s\} = P\{\xi > t\} \tag{43}$$

假如把 ξ 解释为寿命, 则(43)表明, 如果已知寿命长于 s 年, 则再活 t 年的概率与年令 s 无关, 所以有时又风趣地称指数分布是"永远年青"的.

下列事实也是正确的: 指数分布是唯一具有性质(43)的连续型分布.

我们来证明这个论断.

设 ξ 是非负的，其分布函数为 $F(x)$，记

$$G(x)=P\{\xi>x\}$$

则由(43)可以得到

$$G(s+t)=G(s)G(t)$$

对一切 $s\geqslant0, t\geqslant0$ 成立，因为 $G(x)$ 关于 x 单调，所以由第二章 §4 引理知

$$G(x)=a^x,\quad x\geqslant0$$

由于 $G(x)$ 是概率，故 $0<a<1$，可以写成 $a=e^{-\lambda}$，其中 $\lambda>0$. 又因为 $F(x)$ 是连续型分布，故

$$F(x)=1-G(x)=1-e^{-\lambda x},\qquad x\geqslant0$$

从而证明了结论.

应当指出，指数分布与普阿松过程有密切关系，若以 $\xi(t)$ 记参数为 λt 的普阿松过程，以 τ_1 记它第一个跳跃发生的时刻，则

$$P\{\tau_1>t\}=P\{\xi(t)=0\}$$

因此

$$P\{\tau_1>t\}=e^{-\lambda t}$$

这说明 τ_1 服从指数分布(42). 下面把这个结果推广到更为一般的场合.

[Γ-分布] 若 $\xi(t)$ 是参数为 λt 的普阿松过程，以 τ_r 记它的第 r 个跳跃发生的时刻. 事件 $\{\tau_r<t\}$ 发生表明第 r 个跳跃出现在时刻 t 之前，因此事件 $\{\xi(t)\geqslant r\}$ 发生，即 $\{\tau_r<t\}\subset\{\xi(t)\geqslant r\}$；反之，若事件 $\{\xi(t)\geqslant r\}$ 发生，即在时刻 t 时 $\xi(t)$ 之值不小于 r，这时第 r 个跳跃已经出现过，因此事件 $\{\tau_r<t\}$ 发生，即有 $\{\xi(t)\geqslant r\}\subset\{\tau_r<t\}$. 综上所述可知

$$\{\tau_r<t\}=\{\xi(t)\geqslant r\}\tag{44}$$

以 $F(x)$ 记 τ_r 的分布函数，则

$$F(t)=P\{\tau_r<t\}=P\{\xi(t)\geqslant r\}$$

$$=\sum_{k=r}^{\infty}\frac{(\lambda t)^k e^{-\lambda t}}{k!}=1-\sum_{k=0}^{r-1}\frac{(\lambda t)^k e^{-\lambda t}}{k!}$$

因此

$$p(t)=F'(t)$$

$$=-\left[\sum_{k=0}^{r-1}\frac{(\lambda t)^k e^{-\lambda t}(-\lambda)}{k!}+\sum_{k=0}^{r-1}\frac{k(\lambda t)^{k-1}\cdot\lambda e^{-\lambda t}}{k!}\right]$$

$$=\lambda e^{-\lambda t}\sum_{k=0}^{r-1}\frac{(\lambda t)^k}{k!}-\lambda e^{-\lambda t}\sum_{k=1}^{r-1}\frac{(\lambda t)^{k-1}}{(k-1)!}$$

$$=\frac{\lambda(\lambda t)^{r-1}}{(r-1)!}e^{-\lambda t}=\frac{\lambda^r}{\Gamma(r)}t^{r-1}e^{-\lambda t}$$

因对于任意的 $r>0$, $\lambda>0$,

$$\int_0^{\infty}\frac{\lambda^r}{\Gamma(r)}x^{r-1}e^{-\lambda x}dx=1 \tag{45}$$

我们称密度函数为

$$p(x)=\begin{cases}\dfrac{\lambda^r}{\Gamma(r)}x^{r-1}e^{-\lambda x}, & x>0\\ 0, & x\leqslant 0\end{cases} \tag{46}$$

的分布为 **Γ-分布**. 其中 $\lambda>0$, $r>0$ 为参数. 简记作 $G(\lambda, r)$ (图 6).

前面推导说明普阿松过程中第 r 个跳跃发生的时刻服从 Γ-分布. 此外, Γ-分布在概率论, 数理统计, 随机过程中还有不少应用.

显然当 $r=1$ 时, Γ-分布化为指数分布. 另外, 当 r 为正整数时, 服从 Γ-分布的随机变量可以看作是 r 个服从指数分布的随机变量之和, 这个性质与巴斯卡分布类似.

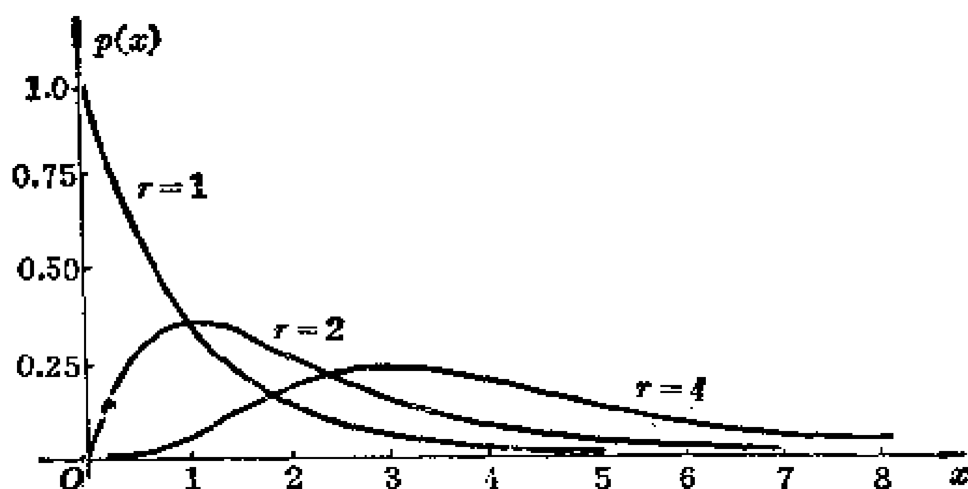


图6 Γ-分布密度函数

*五、关于分布函数的一些结论

我们已经证明了分布函数是单调函数，利用实变数函数论中关于单调函数的一般结果①，不难推出分布函数具有如下性质：

(1) 分布函数至多只有可列个不连续点；

(2) 对分布函数 $F(x)$ 有勒贝格分解

$$F(x)=c_1F_1(x)+c_2F_2(x)+c_3F_3(x) \tag{47}$$

其中 $F_1(x)$ 是跳跃函数，$F_2(x)$ 是绝对连续函数，$F_3(x)$ 是所谓奇异函数，它们都是分布函数；而 $0\leqslant c_i\leqslant 1$, $i=1,2,3$, 且 $c_1+c_2+c_3=1$.

在我们讨论过的分布函数中，离散型分布函数是跳跃函数，相当于在(47)中取 $c_1=1$, $c_2=c_3=0$ 的场合；而连续型分布函数是绝对连续函数，相当于在(47)中取 $c_2=1$, $c_1=c_3=0$ 的场合；自然会想到，也可以取 $c_3=1$, $c_1=c_2=0$, 得到另一类分布函数．这个结论是正确的，理论上确实存在着另一类分布函数——奇异型分布函数，它是连续函数，但却不能表为不定积分，因此它没有密度函数．不过到目前为止，常用的分布都是离散型或连续型的，因此

① 参看复旦大学数学系，《实变数函数论与泛函分析概要》(第二版)，上海科技出版社，1963，第四章．

我们不准备对奇异型分布多加讨论．以后证明结果或是对离散型进行，或是对连续型进行，或者对一般分布进行，不再强调三种类型的混合了．

§2．随机向量，随机变量的独立性

一、随机向量及其分布

在有些随机现象中，每次试验的结果不能只用一个数来描述，而要同时用几个数来描述，例如对于钢的成份，需要同时指出它的含碳量，含硫量，含磷量，等等．这样，对应于每个样本点 ω，试验的结果将是一个向量 $(\xi_1(\omega), \xi_2(\omega), \cdots, \xi_n(\omega))$，这个向量取值于 n 维欧几里得空间 R^n．

定义 若随机变量 $\xi_1(\omega)$, $\xi_2(\omega)$, $\cdots$, $\xi_n(\omega)$ 定义在同一概率空间 $(\Omega, \mathscr{F}, P)$ 上，则称

$$\xi(\omega)=(\xi_1(\omega), \xi_2(\omega), \cdots, \xi_n(\omega)) \tag{1}$$

构成一个 **n 维随机向量**，亦称 **n 维随机变量**．显然，一维随机向量即为随机变量．

固然可以对随机向量的一个个分量分别研究，但是我们马上就会看到，把它们作为一个向量，则不但能研究各个分量的性质，而且还可以考察它们之间的联系，对许多问题来说，这是十分必要的．

对于 n 个实数 $x_1, x_2, \cdots, x_n$，

$$\{\xi_1(\omega)<x_1, \xi_2(\omega)<x_2, \cdots, \xi_n(\omega)<x_n\}$$

$$=\bigcap_{i=1}^{n}\{\xi_i(\omega)<x_i\}\in\mathscr{F} \tag{2}$$

亦即对于 R^n 中的 n 维区间 $C_n=\prod_{i=1}^{n}(-\infty, x_i)$，有

$$\{\xi(\omega)\in C_n\}\in\mathscr{F}$$

利用测度论的方法还可证明，若 B_n 为 R^n 上任一波雷尔点集，也有

$$\{\xi(\omega)\in B_n\}\in\mathscr{F} \tag{3}$$

以后我们将要用到这个结论(图 1)．

类似于一维的场合，我们引进如下定义．

定义 称 n 元函数

$$F(x_1, x_2, \cdots, x_n)=P(\xi_1(\omega)<x_1, \xi_2(\omega)<x_2, \cdots, \xi_n(\omega)<x_n\} \tag{4}$$

为随机向量 $\xi(\omega)=(\xi_1(\omega), \xi_2(\omega), \cdots, \xi_n(\omega))$ 的（联合）分布函数．

给定了联合分布函数后，可以计算事件$\{a_1\leqslant\xi_1<b_1, a_2\leqslant\xi_2<b_2, \cdots, a_n\leqslant\xi_n<b_n\}$的概率，例如当 $n=2$ 时，

$$P\{a_1\leqslant\xi_1<b_1, a_2\leqslant\xi_2<b_2\}$$
$$=F(b_1, b_2)-F(a_1, b_2)-F(b_1, a_2)+F(a_1, a_2) \tag{5}$$

这个结果容易从图 2 看出．

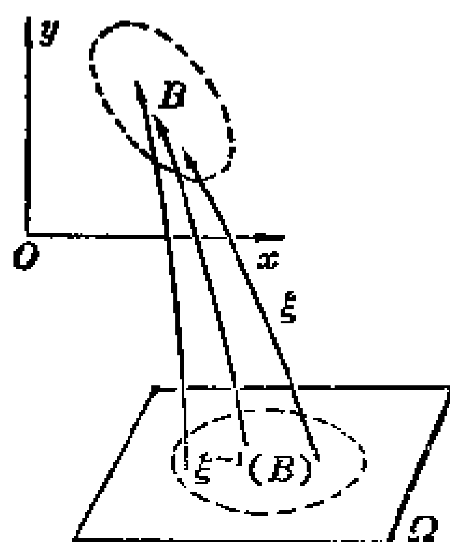


图 1 二维随机向量

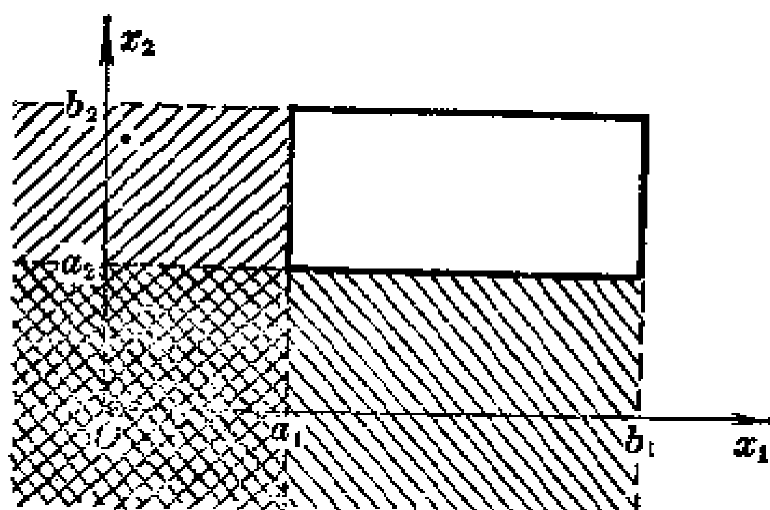


图 2 二维概率计算

类似于一维的场合，可以证明多元分布函数的一些性质．例如关于每个变元单调，关于每个变元左连续，以及

$$F(x_1, x_2, \cdots, -\infty, \cdots, x_n)=0 \tag{6}$$

$$F(+\infty, +\infty, \cdots, +\infty)=1 \tag{7}$$

此外，利用概率的非负性，可以知道(5)式右边是非负的．反之，满足这些要求的二元函数是某随机变量的分布函数．类似的结论对 n 维场合显然也成立．

随机向量的分布函数也有离散型与连续型的分别．在离散型场合，概率分布集中在有限或可列个点上，例如第二章 §3 中给出的多项分布，就是一个例子．在连续型场合，存在着非负函数 $p(x_1, x_2, \cdots, x_n)$，使

$$F(x_1, \cdots, x_n)=\int_{-\infty}^{x_1}\cdots\int_{-\infty}^{x_n} p(y_1, \cdots, y_n)dy_1\cdots dy_n \tag{8}$$

这里的 $p(x_1, \cdots, x_n)$称为密度函数，满足如下两个条件

$$p(x_1, \cdots, x_n)\geqslant 0 \tag{9}$$

$$\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} p(x_1, \cdots, x_n)dx_1\cdots dx_n=1 \tag{10}$$

随机向量的概念，在各个基础学科和工程技术中已有广泛应用．例如在量子力学中，粒子在某个区域 G 的出现是通过概率来描述的．若以 ψ 表示它的波函数，则 $|\psi|^2$ 即为密度函数，而

$$\iiint_G |\psi|^2 dxdydz$$

就给出该粒子在区域 G 出现的概率．

均匀分布和 n 元正态分布是比较常见的两种多维连续型分布．

［均匀分布］ 若 G 为 R^n 中有限区域，其测度 $S>0$；则由密度函数

$$p(x_1, \cdots, x_n)=\begin{cases}\dfrac{1}{S}, & (x_1, \cdots, x_n)\in G \\ 0, & (x_1, \cdots, x_n)\overline{\in} G\end{cases} \tag{11}$$

给出的分布称为**G上的均匀分布**.

在第一章几何概率一节中，我们已看到均匀分布的各种例子.

[多元正态分布] 若 $B=(b_{ij})$是 n 阶正定对称矩阵，以 $B^{-1}=(r_{ij})$表示 B 的逆阵；$|B|$表示 B 的行列式的值. $a=(a_1,\cdots,a_n)$是任意实值行向量，则由密度函数

$$p(x_1,\cdots,x_n)=\frac{1}{(2\pi)^{\frac{n}{2}}|B|^{\frac{1}{2}}}\exp\left\{-\frac{1}{2}\sum_{j,k=1}^{n}r_{jk}(x_j-a_j)(x_k-a_k)\right\} \tag{12}$$

定义的分布称为**n元正态分布**，简记为 $N(a,B)$.

这个密度函数也可以改为如下向量形式：

$$p(x)=\frac{1}{(2\pi)^{\frac{n}{2}}|B|^{\frac{1}{2}}}\exp\left\{-\frac{1}{2}(x-a)B^{-1}(x-a)'\right\} \tag{13}$$

这里$(x-a)'$ 表示向量$(x-a)$的转置.

n 元正态分布是最重要的一种多维分布，它在概率论，数理统计，随机过程论中都占有重要地位，具有许多重要性质. 对于这些性质的叙述和证明，我们将在引进了更有力的工具后再进行. 这里我们将对它的特殊场合——二元正态分布逐步直接导出这些性质.

二、边际分布

为方便起见，讨论将对二维场合进行，多维时这些结论仍然成立.

先讨论离散型分布的场合，在这种场合，有关概念特别容易理解.

考虑二维随机向量(ξ,η)，设 ξ 取值 $x_1,x_2,\cdots$；η 取值 $y_1,y_2,\cdots$. 记

$$P\{\xi=x_i,\eta=y_j\}=p(x_i,y_j),\quad i,j=1,2,\cdots \tag{14}$$

$$P\{\xi=x_i\}=p_1(x_i),\qquad i=1,2,\cdots \tag{15}$$

$$P\{\eta=y_j\}=p_2(y_j),\qquad j=1,2,\cdots \tag{16}$$

显然

$$p(x_i,y_j)\geqslant 0,\quad \sum_{i,j}p(x_i,y_j)=1 \tag{17}$$

此外，对固定的 i,

$$\sum_j p(x_i,y_j)=P\{\xi=x_i\}=p_1(x_i) \tag{18}$$

而对固定的 j

$$\sum_i p(x_i,y_j)=P\{\eta=y_j\}=p_2(y_j) \tag{19}$$

换句话说，由联合概率分布，对于固定的 i 关于 j 求和得到 ξ 的概率分布；而对于固定的 j 关于 i 求和得到 η 的概率分布.

这里 $p_1(x_i)$ 与 $p_2(y_j)$ 称为 $p(x_i,y_j)$ 的**边际分布**. 这个名称的含义通过下列两例将看得很清楚.

［例 1］ 袋中装有 2 只白球及 3 只黑球. 现进行有放回的摸球，定义下列随机变量

$$\xi=\begin{cases}1, & \text{第 1 次摸出白球}\\ 0, & \text{第 1 次摸出黑球}\end{cases}\qquad \eta=\begin{cases}1, & \text{第 2 次摸出白球}\\ 0, & \text{第 2 次摸出黑球}\end{cases}$$

则 (ξ,η) 的联合概率分布与边际分布由表 1 给出.

［例 2］ 前例中若采用不放回摸球，则 (ξ,η) 的联合概率分布及边际分布由表 2 给出.

表 1 有放回摸球的概率分布

η \ ξ	0	1	$p_2(y_j)$
0	$\frac{3}{5}\cdot\frac{3}{5}$	$\frac{2}{5}\cdot\frac{3}{5}$	$\frac{3}{5}$
1	$\frac{3}{5}\cdot\frac{2}{5}$	$\frac{2}{5}\cdot\frac{2}{5}$	$\frac{2}{5}$
$p_1(x_i)$	$\frac{3}{5}$	$\frac{2}{5}$	

表 2 不放回摸球的概率分布

η \ ξ	0	1	$p_2(y_j)$
0	$\frac{3}{5}\cdot\frac{2}{4}$	$\frac{2}{5}\cdot\frac{3}{4}$	$\frac{3}{5}$
1	$\frac{3}{5}\cdot\frac{2}{4}$	$\frac{2}{5}\cdot\frac{1}{4}$	$\frac{2}{5}$
$p_1(x_i)$	$\frac{3}{5}$	$\frac{2}{5}$	

在上面两个表中，中间部分是(ξ,η)的联合概率分布，而边沿部分是ξ及η的概率分布，它们由联合分布经同一行或同一列的相加而得出来，“边际”二字即由上面双行表的外貌而来.

让我们再注意一个重要事实，两例中ξ及η的(边际)分布是相同的，但是它们的联合分布却完全不同．这里可以看出联合分布不能由边际分布唯一确定，也就是说二维随机向量的性质并不能由它两个分量的个别性质来确定，这时还必须考虑它们之间的联系，这也说明了研究多维随机向量的作用.

一般地，若(ξ,η)是二维随机向量，其分布函数为$F(x,y)$，我们能由$F(x,y)$得出ξ或η的分布函数.事实上，

$$F_1(x)=P\{\xi<x\}=P\{\xi<x,\eta<\infty\}=F(x,+\infty) \tag{20}$$

同理

$$F_2(y)=P\{\eta<y\}=F(+\infty,y)$$

$F_1(x)$及$F_2(y)$称为$F(x,y)$的**边际分布函数**.

若$F(x,y)$是连续型分布函数，有密度函数$p(x,y)$，那么

$$F_1(x)=\int_{-\infty}^{x}\int_{-\infty}^{\infty}p(u,y)dudy$$

因此$F_1(x)$是连续型分布函数，其密度函数为

$$p_1(x)=\int_{-\infty}^{\infty}p(x,y)dy \tag{21}$$

同理$F_2(x)$是连续型分布函数，其密度函数为

$$p_2(y)=\int_{-\infty}^{\infty}p(x,y)dx$$

$p_1(x)$及$p_2(y)$称为$p(x,y)$的**边际分布密度函数**.

[二元正态分布]　函数

$$p(x,y)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}}\exp\left\{-\frac{1}{2(1-r^2)}\cdot\left[\frac{(x-a)^2}{\sigma_1^2}-\frac{2r(x-a)(y-b)}{\sigma_1\sigma_2}+\frac{(y-b)^2}{\sigma_2^2}\right]\right\} \tag{22}$$

这里 $a, b, \sigma_1, \sigma_2, r$ 为常数，$\sigma_1>0, \sigma_2>0, |r|<1$，称为**二元正态分布密度函数**(图 3)。

显然这是 n 元正态分布当 $n=2$ 时的特殊情况，相应的

$$B=\begin{pmatrix}\sigma_1^2 & r\sigma_1\sigma_2\\ r\sigma_1\sigma_2 & \sigma_2^2\end{pmatrix},\quad a_1=a,\quad a_2=b$$

我们来求其边际分布密度函数，令 $\frac{x-a}{\sigma_1}=u, \frac{y-b}{\sigma_2}=v$，由(21)知

$$\begin{aligned}p_1(x)&=\int_{-\infty}^{\infty}p(x,y)dy\\
&=\frac{1}{2\pi\sigma_1\sqrt{1-r^2}}\int_{-\infty}^{\infty}\exp\left\{-\frac{1}{2(1-r^2)}\cdot[u^2-2ruv+v^2]\right\}dv\\
&=\frac{1}{\sqrt{2\pi}\,\sigma_1}e^{-u^2/2}\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi(1-r^2)}}\\
&\quad\cdot\exp\left\{-\frac{r^2u^2-2ruv+v^2}{2(1-r^2)}\right\}dv\\
&=\frac{1}{\sqrt{2\pi}\sigma_1}e^{-u^2/2}\cdot\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi(1-r^2)}}e^{-(v-ru)^2/2(1-r^2)}dv\\
&=\frac{1}{\sqrt{2\pi}\ \sigma_1}e^{-u^2/2}\\
&=\frac{1}{\sqrt{2\pi}\ \sigma_1}e^{-(x-a)^2/2\sigma_1^2}\end{aligned}\tag{23}$$

即 $p_1(x)$ 是 $N(a,\sigma_1^2)$ 的密度函数。同理

$$p_2(y)=\frac{1}{\sqrt{2\pi}\ \sigma_2}e^{-(y-b)^2/2\sigma_2^2}\tag{24}$$

因此二元正态分布的边际分布仍为正态分布，这是一个重要的结论。

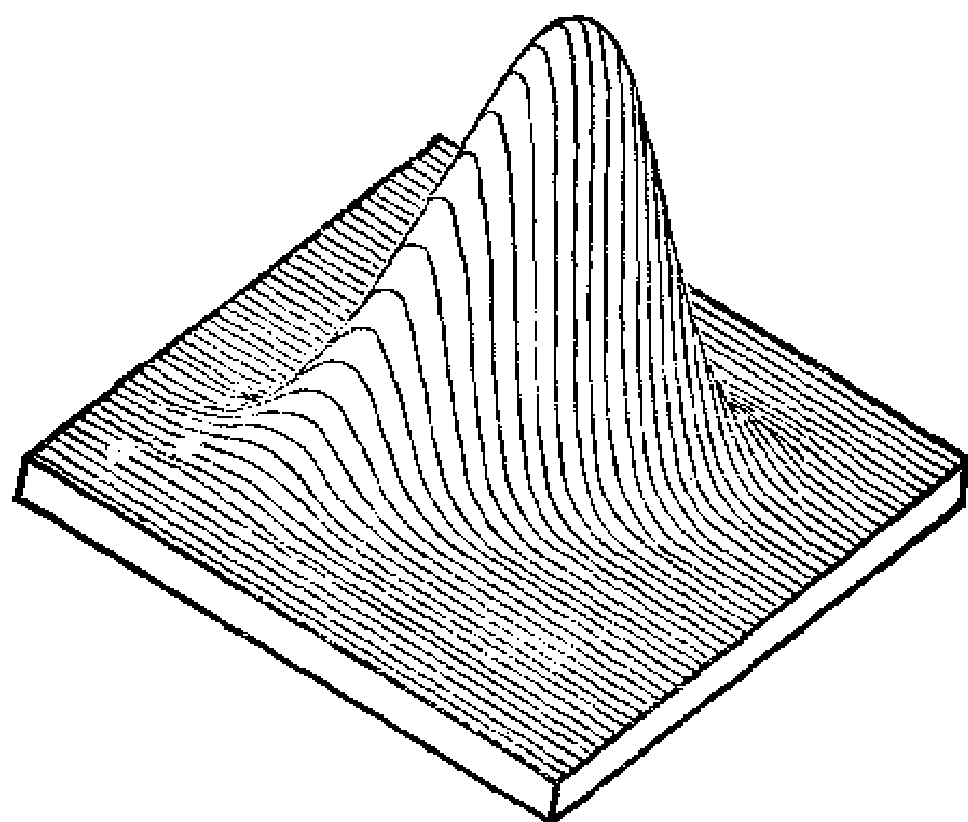


图3　二维正态密度曲面

由(22)定义的 $p(x,y)$ 显然非负，又由于

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}p(x,y)dxdy=\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}\sigma_1}e^{-(x-a)^2/2\sigma_1^2}dx=1$$

因此我们已顺带验证了 $p(x,y)$ 确实是一个密度函数。

三、条件分布

对于多个随机事件可以讨论它们的条件概率，同样地，对于多个随机变量也可以讨论它们的条件分布，并由此得出重要结果。

仍对二维的场合进行讨论。也还是先从离散型开始，这时并无多大困难。

若已知 $\xi=x_i$，$(p_1(x_i)>0)$ 则事件 $\{\eta=y_j\}$ 的条件概率为

$$P\{\eta=y_j|\xi=x_i\}=\frac{P\{\xi=x_i,\eta=y_j\}}{P\{\xi=x_i\}}=\frac{p(x_i,y_j)}{p_1(x_i)} \tag{25}$$

这式子定义了随机变量 η 关于随机变量 ξ 的**条件分布**。在一般情况下，它不同于 $p_2(y_j)$，这表示从 ξ 的值可以得出关于 η 的某

些结论，反之亦然.

对于一般随机向量 (ξ, η)，我们也想定义条件分布函数 $P\{\eta<y|\xi=x\}$，但是由于会出现 $P\{\xi=x\}=0$，因此我们不能象(25)那样简单地定义.

自然会想到可以用

$$
\begin{aligned}
P\{\eta<y|\xi=x\} &= \lim_{\Delta x\to 0} P\{\eta<y|x\leqslant\xi<x+\Delta x\} \\
&= \lim_{\Delta x\to 0}\frac{P\{x\leqslant\xi<x+\Delta x, \eta<y\}}{P\{x\leqslant\xi<x+\Delta x\}} \\
&= \lim_{\Delta x\to 0}\frac{F(x+\Delta x, y)-F(x, y)}{F(x+\Delta x, \infty)-F(x, \infty)} \qquad (26)
\end{aligned}
$$

来定义.

特别是对于有连续密度函数的场合，这定义导出

$$
P\{\eta<y|\xi=x\}=\lim_{\Delta x\to 0}\frac{\int_x^{x+\Delta x}\int_{-\infty}^{y} p(u, v)dudv}{\int_x^{x+\Delta x}\int_{-\infty}^{\infty} p(u, v)dudv}
$$

利用中值定理，当 $p_1(x)\neq 0$ 时，

$$
P\{\eta<y|\xi=x\}=\frac{\int_{-\infty}^{y} p(x, v)dv}{p_1(x)}=\int_{-\infty}^{y}\frac{p(x, v)}{p_1(x)}dv
$$

因此在给定 $\xi=x$ 的条件下，η 的分布密度函数为

$$
p(y|x)=\frac{p(x, y)}{p_1(x)} \qquad (27)
$$

同理可得在给定 $\eta=y$ 的条件下，ξ 的分布密度函数为

$$
p(x|y)=\frac{p(x, y)}{p_2(y)} \qquad (28)
$$

这里当然也要求 $p_2(y)\neq 0$.

［例 3］ 对由 (22) 定义的二元正态分布求条件密度函数 $p(y|x)$，由(27)

$$p(y|x)=\frac{p(x,y)}{p_1(x)}=\frac{1}{\sigma_2\sqrt{2\pi}\sqrt{1-r^2}}\exp\left\{-\frac{1}{2(1-r^2)}\right.$$

$$\left.\cdot\left[\frac{(x-a)^2}{\sigma_1^2}-\frac{2r(x-a)(y-b)}{\sigma_1\sigma_2}+\frac{(y-b)^2}{\sigma_2^2}\right]+\frac{(x-a)^2}{2\sigma_1^2}\right\}$$

$$=\frac{1}{\sigma_2\sqrt{2\pi}\sqrt{1-r^2}}\exp\left\{-\frac{1}{2(1-r^2)}\right.$$

$$\left.\cdot\left[\frac{(x-a)^2r^2}{\sigma_1^2}-\frac{2r(x-a)(y-b)}{\sigma_1\sigma_2}+\frac{(y-b)^2}{\sigma_2^2}\right]\right\}$$

$$=\frac{1}{\sigma_2\sqrt{2\pi}\sqrt{1-r^2}}\exp\left\{-\frac{1}{2(1-r^2)}\left(\frac{y-b}{\sigma_2}-r\frac{x-a}{\sigma_1}\right)^2\right\}$$

$$=\frac{1}{\sqrt{2\pi}\,\sigma_2\sqrt{1-r^2}}\exp\left\{-\frac{1}{2\sigma_2^2(1-r^2)}\right.$$

$$\left.\cdot\left[y-\left(b+r\frac{\sigma_2}{\sigma_1}(x-a)\right)\right]^2\right\} \tag{29}$$

从这里我们看到，二元正态分布的条件分布仍然是正态分布

$$N\left(b+r\frac{\sigma_2}{\sigma_1}(x-a),\ \sigma_2^2(1-r^2)\right) \tag{30}$$

特别指出，这里 $m=b+r\frac{\sigma_2}{\sigma_1}(x-a)$ 是 x 的线性函数，这个结论在一些统计问题中很重要.

四、随机变量的独立性

在上章中，我们看到随机事件的独立性起着很大的作用. 下面研究随机变量的独立性. 引入如下定义.

定义 设 $\xi_1,\cdots,\xi_n$ 为 n 个随机变量，若对于任意的 $x_1,\cdots,x_n$ 成立

$$P\{\xi_1<x_1,\cdots,\xi_n<x_n\}=P\{\xi_1<x_1\}\cdots P\{\xi_n<x_n\} \tag{31}$$

则称 $\xi_1,\cdots,\xi_n$ 是相互独立的.

若 ξ_i 的分布函数为 $F_i(x)$，它们的联合分布函数为 $F(x_1,\cdots,x_n)$，则(31)等价于对一切 $x_1,\cdots,x_n$ 成立

$$F(x_1,\cdots,x_n)=F_1(x_1)\cdots F_n(x_n) \tag{32}$$

在这种场合，由每个随机变量的(边际)分布函数可以唯一地确定联合分布函数．而且由(26)可以看到，这时条件分布化为无条件分布

$$P\{\eta<y\mid\xi=x\}=P\{\eta<y\} \tag{33}$$

即由 ξ 的值不能得出任何关于 η 的结论．

对于离散型随机变量，(31) 等价于对任何一组可能取的值 $(x_1,\cdots,x_n)$ 成立

$$P\{\xi_1=x_1,\cdots,\xi_n=x_n\}=P\{\xi_1=x_1\}\cdots P\{\xi_n=x_n\} \tag{34}$$

例 1 中有放回摸球时的 ξ 与 η 是相互独立的，这时联合分布取乘法表的形式；反之，例 2 中的 ξ 与 η 不是独立的．这些例子也说明，如果两个随机变量有相同的分布，则它们可以是独立的，也可以不是独立的．

对于连续型随机变量，条件(31)的等价形式是对一切 $x_1,\cdots,x_n$ 成立

$$p(x_1,\cdots,x_n)=p_1(x_1)\cdots p_n(x_n) \tag{35}$$

这里 $p(x_1,\cdots,x_n)$是联合分布密度函数，而 $p_i(x_i)$ 是各随机变量的密度函数．

［例 4］ 对由(22)定义的二元正态分布，有

$$p_1(x)p_2(y)=\frac{1}{2\pi\sigma_1\sigma_2}\exp\left\{-\frac{(x-a)^2}{2\sigma_1^2}-\frac{(y-b)^2}{2\sigma_2^2}\right\}$$

与(22)比较可知，要关系式(35)成立的充要条件是

$$r=0 \tag{36}$$

即服从二元正态分布的随机变量独立的充要条件是 $r=0$，这时条件分布(29)化为

$$p(y\mid x)=\frac{1}{\sqrt{2\pi}\,\sigma_2}e^{-(y-b)^2/2\sigma_2^2}=p_2(y)$$

这与(33)是一致的.

上面我们列举了随机变量独立性的各种表达形式，有些是对一般随机变量成立的，有些只对离散型或连续型才成立，一般说来，这些条件比较易于验证，下面我们介绍另外一个条件，这条件不易验证，但在理论研究中有用.

随机变量 $\xi_1, \xi_2, \cdots, \xi_n$ 相互独立的充要条件是对一切一维波雷尔点集 $B_1, B_2, \cdots, B_n$ 成立

$$P\{\xi_1\in B_1, \xi_2\in B_2, \cdots, \xi_n\in B_n\}$$
$$=P\{\xi_1\in B_1\}P\{\xi_2\in B_2\}\cdots P\{\xi_n\in B_n\} \tag{37}$$

论断的证明要用到测度论，已超出本课程范围.

当然也可以建立 n 维随机向量 ξ 与 m 维随机向量 η 相互独立的概念，这时要求成立

$$P\{\xi\in A, \eta\in B\}=P\{\xi\in A\}P\{\eta\in B\}$$

其中 A, B 分别是任意一个 n 维及 m 维的波雷尔点集.

显然若 ξ 与 η 独立，则 ξ 的子向量与 η 的子向量是独立的.

此外，注意到若 $\xi_1, \xi_2, \cdots, \xi_n$ 相互独立，则其中的任意 $r(2\leqslant r<n)$ 个随机变量也相互独立. 例如，我们证明 $\xi_1, \xi_2, \cdots, \xi_{n-1}$ 相互独立.

$$P\{\xi_1<x_1, \cdots, \xi_{n-1}<x_{n-1}\}=P\{\xi_1<x_1, \cdots, \xi_{n-1}<x_{n-1}, \xi_n<\infty\}$$
$$=P\{\xi_1<x_1\}\cdots P\{\xi_{n-1}<x_{n-1}\}P\{\xi_n<\infty\}$$
$$=P\{\xi_1<x_1\}\cdots P\{\xi_{n-1}<x_{n-1}\}$$

随机变量的独立性概念是概率论中最基本的概念之一，也是最重要的概念之一，关于独立随机变量的研究构成了概率论的重要课题，我们将在第五章中介绍一些基本结果.

有了独立性的概念，可以介绍一个导出正态分布的条件. 先证明一个分析引理.

引理 若 $f(x)$ 及 $g(y)$ $(x\geqslant 0,\ y\geqslant 0)$ 是不恒等于 0 的连续函

数，而且对一切非负实数 x 及 y 成立

$$f(x)g(y)=h(x+y) \tag{38}$$

则

$$f(x)=ka^x,\quad x\geqslant 0 \tag{39}$$

这里 k 及 a 是常数，$a>0$.

［证明］ 首先证明 $f(0)\neq 0$. 用反证法，假定 $f(0)=0$，则 $h(y)\equiv 0$，由于 $g(y)$不恒为 0，有一个 y_0 使 $g(y_0)\neq 0$. 因此 $f(x)=\dfrac{h(x+y_0)}{g(y_0)}\equiv 0$，这与 $f(x)$不恒为 0 的条件矛盾，因此 $f(0)\neq 0$. 同理可证 $g(0)\neq 0$.

其次在(38)中令 $x=z, y=0$，得 $f(z)g(0)=h(z)$，又在(38)中令 $x=0, y=z$ 得 $f(0)g(z)=h(z)$，因此

$$f(z)g(0)=f(0)g(z)=h(z)$$

两边除以 $f(0)g(0)$并记得到的比值为 $p(z)$，则得

$$\frac{f(z)}{f(0)}=\frac{g(z)}{g(0)}=\frac{h(z)}{f(0)g(0)}=p(z) \tag{40}$$

代入(38)得

$$p(x)p(y)=p(x+y) \tag{41}$$

利用第二章 § 4 引理知

$$p(x)=a^x$$

因此

$$f(x)=f(0)a^x$$

引理得证.

［射击弹落点分布］ 在火器射击问题中，假如过射击目标中心作直角坐标轴 x 轴及 y 轴，每次射击弹落点由于随机因素影响会偏离目标，它的坐标(ξ,η)是一个二维随机变量. 假定它满足下列条件

(1) ξ 与 η 具有连续的密度函数，分别记为 $p_1(x)$ 及 $p_2(y)$；

(2) ξ 与 η 相互独立；

(3) (ξ, η) 的密度函数在 (x, y) 点的值仅与它到原点的距离 $r=\sqrt{x^2+y^2}$ 有关。则 ξ 与 η 均服从正态分布。

[证明] 记 (ξ, η) 的密度函数为 $p(x, y)$，由假定(3)知 $p(x, y)=q(\sqrt{x^2+y^2})$，再由假定(1)，(2)知

$$p_1(x)p_2(y)=q(\sqrt{x^2+y^2}) \tag{42}$$

因为 $p_1(x)$ 及 $p_2(y)$ 是密度函数，当然不能恒等于0，故可利用引理[把 $p_1(x)$ 看作 $f(x^2)$，$p_2(y)$ 看作 $g(y^2)$，$q(\sqrt{x^2+y^2})$ 看作 $h(x^2+y^2)$]，得到

$$p_1(x)=ka^{x^2}$$

由于 $p_1(x)$ 是 $(-\infty, \infty)$ 上的密度函数，因此必须有 $0<a<1$，记 $a=e^{-1/2\sigma_1^2}$，其中 $\sigma_1>0$，利用 $\int_{-\infty}^{\infty} p_1(x)dx=1$，可知 $k=\frac{1}{\sqrt{2\pi}\,\sigma_1}$，这样一来，

$$p_1(x)=\frac{1}{\sqrt{2\pi}\,\sigma_1}e^{-x^2/2\sigma_1^2}$$

因此 ξ 服从正态分布. 同理可证 η 也服从正态分布.

§3. 随机变量的函数及其分布

一、波雷尔函数与随机变量的函数

在许多问题中需要计算随机变量的函数的分布律，例如在无线电接收中，某时刻收到的信号是一个随机变量 ξ，若我们把这个信号通过平方检波器，则输出的信号为 $\eta=\xi^2$，这时我们需要求 η 的分布律. 又如在统计物理中，已知分子运动速度 ξ 的分布，要求其动能 $\eta=\frac{1}{2}m\xi^2$ 的分布律. 这类问题既普遍而又重要，接下来我

们就要讨论它.

这类问题较为一般的提法是：若 ξ 是随机变量，求 $\eta=g(\xi)$ 的分布律.

为了使 η 有分布律可言，当然要求 η 是随机变量，因此对函数 $y=g(x)$ 也必须有一定的要求，为此我们提出如下定义.

定义 设 $y=g(x)$ 是 R^1 到 R^1 上的一个映照，若对于一切 R^1 中的波雷尔点集 B_1 均有

$$\{x:g(x)\in B_1\}\in\mathscr{B}_1 \tag{1}$$

其中 $\mathscr{B}_1$ 为 R^1 上波雷尔 σ-域，则称 $g(x)$ 是**一元波雷尔（可测）函数**.

波雷尔函数是很广泛的一类函数，我们所碰到的大部分函数都是波雷尔函数，特别地，已知所有的连续函数及单调函数都是波雷尔函数.

若 ξ 是概率空间 $(\Omega,\ \mathscr{F},\ P)$ 上的随机变量，而 $g(x)$ 是一元波雷尔函数，则 $g(\xi)$ 是 $(\Omega,\ \mathscr{F},P)$ 上的随机变量. 事实上，对一切 $B_1\in\mathscr{B}_1$ 有

$$\{\omega:g(\xi(\omega))\in B_1\}=\{\omega:\xi(\omega)\in g^{-1}(B_1)\}\in\mathscr{F} \tag{2}$$

这里 $g^{-1}(B_1)=\{x:g(x)\in B_1\}$，由(1)知它是一维波雷尔点集，再根据随机变量定义中的(1.1)式即得(2).

有时，我们还要考虑随机向量的函数，例如 $\eta=\xi_1+\xi_2$ 就是随机向量 (ξ_1,ξ_2) 的函数. 一般地，要研究 $\eta=g(\xi_1,\ \cdots,\ \xi_n)$ 及其分布. 这时对 n 元函数 $y=g(x_1,\cdots,x_n)$ 也要有相应的要求.

定义 设 $y=g(x_1,\ \cdots,\ x_n)$ 是 R^n 到 R^1 上的一个映照，若对一切 R^1 中的波雷尔点集 B_1 均有

$$\{(x_1,\cdots,x_n):g(x_1,\cdots,x_n)\in B_1\}\in\mathscr{B}_n \tag{3}$$

其中 $\mathscr{B}_n$ 为 R^n 上波雷尔 σ-域，则称 $g(x_1,\ \cdots,\ x_n)$ 为 **n 元波雷尔（可测）函数**.

同样地，若$(\xi_1, \cdots, \xi_n)$是$(\Omega, \mathscr{F}, P)$上的随机向量，而$g(x_1, \cdots, x_n)$是n元波雷尔函数，则$g(\xi_1, \cdots, \xi_n)$是$(\Omega, \mathscr{F}, P)$上的随机变量.

这个事实证明如下：

$$\{\omega: g(\xi_1(\omega), \cdots, \xi_n(\omega)) \in B_1\}$$
$$= \{\omega: (\xi_1(\omega), \cdots, \xi_n(\omega)) \in g^{-1}(B_1)\} \in \mathscr{F} \tag{4}$$

其中$g^{-1}(B_1)$是(3)中点集，它是n维波雷尔点集，而按(2, 3)即得最后的关系式.

更一般地，还要研究n维随机向量$(\xi_1, \cdots, \xi_n)$的m个函数$g_1(\xi_1, \cdots, \xi_n), \cdots, g_m(\xi_1, \cdots, \xi_n)$，这里$g_1, \cdots, g_m$都是$n$元波雷尔函数，这是一个$m$维随机向量，需要求出它们的分布.

一般地来说，对于离散型随机变量，求它的函数的分布并不很困难. 例如：若ξ的分布列为

$$\begin{pmatrix} x_1, x_2, \cdots, x_n, \cdots \\ p_1, p_2, \cdots, p_n, \cdots \end{pmatrix}$$

则$\varphi(\xi)$的分布列可由下法得到，列出

$$\begin{pmatrix} \varphi(x_1), & \varphi(x_2), \cdots, & \varphi(x_n), \cdots \\ p_1, & p_2, \cdots, & p_n, \cdots \end{pmatrix}$$

当然这里可能有某些$\varphi(x_i)$相等，把它们作适当并项即可.

[离散褶积公式] 若ξ与η是相互独立的随机变量，它们都取非负整数值，其概率分布分别为$\{a_k\}$及$\{b_k\}$，下面我们来计算随机变量$\zeta = \xi + \eta$的概率分布. 因为

$$\{\zeta = r\} = \{\xi = 0, \eta = r\} + \{\xi = 1, \eta = r-1\} + \cdots + \{\xi = r, \eta = 0\}$$

利用独立性的假定得到

$$c_r = P\{\zeta = r\} = a_0 b_r + a_1 b_{r-1} + \cdots + a_r b_0, \quad r = 0, 1, 2, \cdots \tag{5}$$

这就是求独立随机变量和的分布的公式——**离散褶积公式**，亦称**离散卷积公式**.

下面的讨论主要对连续型随机变量进行. 我们将由简到繁逐步深入.

二、单个随机变量的函数的分布律

先从两个简单的例子开始.

[例 1] 若随机变量 ξ 有密度函数 $f(x)$, 而 $\eta=a\xi+b$, 这里 $a\neq0$. 求 η 的密度函数 $g(y)$.

[解] 分别记 ξ 与 η 的分布函数为 $F(x)$ 及 $G(y)$, 显然有

$$G(y)=P\{\eta<y\}=P\{a\xi+b<y\}$$

$$=\begin{cases} P\left\{\xi<\dfrac{y-b}{a}\right\}=F\left(\dfrac{y-b}{a}\right), & 若\ a>0 \\ P\left\{\xi>\dfrac{y-b}{a}\right\}=1-F\left(\dfrac{y-b}{a}\right), & 若\ a<0 \end{cases} \tag{6}$$

因此

$$g(y)=\begin{cases} \dfrac{1}{a}f\left(\dfrac{y-b}{a}\right), & 若\ a>0 \\ -\dfrac{1}{a}f\left(\dfrac{y-b}{a}\right), & 若\ a<0 \end{cases} \tag{7}$$

或统一起来, 写成

$$g(y)=\frac{1}{|a|}f\left(\frac{y-b}{a}\right) \tag{8}$$

把例 1 的结果应用到正态分布的场合. 若 ξ 服从 $N(a,\ \sigma^2)$, 则 $\eta=\dfrac{\xi-a}{\sigma}$ 的密度函数

$$g(y)=\sigma\cdot\frac{1}{\sqrt{2\pi}\,\sigma}\exp\left\{-\frac{\left[\sigma\left(y+\dfrac{a}{\sigma}\right)-a\right]^2}{2\sigma^2}\right\}=\frac{1}{\sqrt{2\pi}}e^{-y^2/2}$$

即 η 服从 $N(0,1)$, 这是我们早已提到过的.

[例 2] 若 ξ 为连续型随机变量, 其分布密度为 $p(x)$, 求 $\eta=$ 的分布.

$g(y)=\frac{1}{|\cos y|}f(\arcsin y)$

[解] 记 $G(y)=P\{\sin\xi<y\}$. 当 $y\leqslant -1$ 时, $G(y)=0$, 当 $y>1$ 时, $G(y)=1$. 当 $-1<y\leqslant 1$ 时(图 1),

$$G(y)=P\{\sin\xi<y\}$$

$$=\sum_{k=-\infty}^{\infty}P\{2k\pi-(\pi+\sin^{-1}y)\leqslant\xi<2k\pi+\sin^{-1}y\}$$

$$=\int_{\sin x<y}p(x)dx \tag{9}$$

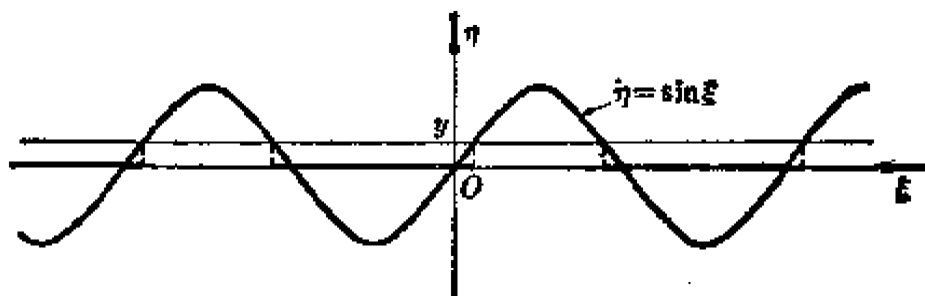


图1 $\eta=\sin\xi$的分布

一般地, 若 ξ 为连续型随机变量, 其密度函数为 $p(x)$, 而 $\eta=f(\xi)$, 我们对下面两种情形进行讨论.

(1) 若 $f(x)$严格单调, 其反函数 $f^{-1}(y)$ 有连续导函数, 则 $\eta=f(\xi)$是具有密度函数①

$$p[f^{-1}(y)]\,|[f^{-1}(y)]'| \tag{10}$$

的连续型随机变量.

[证明] 对于任一实数 a, 记使 $f(x)<a$ 成立的 x 的值的范围为 $E(a)$, 则

$$P\{\eta<a\}=P\{f(\xi)<a\}=P\{\xi\in E(a)\}$$

$$=\int_{E(a)}p(x)dx=\int_{-\infty}^{a}p[f^{-1}(y)]\,|[f^{-1}(y)]'|dy$$

这里用了积分的变数代换. 由上式知道 η 为连续型随机变量, 其密度函数为(10).

① 这里和下面都约定, 对使反函数无意义的 y, 密度函数定义为 0.

利用这个结果直接计算例1中的密度函数. 下面是另

［例 3］ 若 θ 服从$\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$的均匀分布, $\psi=\operatorname{tg}\theta$, 试求 ψ 的密度函数 $g(y)$.

［解］ 记 $y=\operatorname{tg} x$, 则 $x=\operatorname{tg}^{-1}y, \frac{dx}{dy}=\frac{1}{1+y^2}$, 因此由(10)知

$$g(y)=\frac{1}{\pi}\cdot\frac{1}{1+y^2}, \quad -\infty<y<\infty \tag{11}$$

(11)定义的分布称为**柯西分布**, 它是概率论中有名的分布之一.

(2) 若 $f(x)$ 在不相重叠的区间 $I_1, I_2, \cdots$ 上逐段严格单调, 其反函数分别为 $h_1(y), h_2(y), \cdots$, 而且 $h_1'(y), h_2'(y), \cdots$ 均为连续函数, 那么 $\eta=f(\xi)$ 是连续型随机变量, 其密度函数为

$$p[h_1(y)]|h_1'(y)|+p[h_2(y)]|h_2'(y)|+\cdots \tag{12}$$

［证明］ 给定实数 a, 以 $E_i(a)$ 记 I_i 中使 $f(x)<a$ 成立的 x 的集合, 显然诸 $E_i(a)$ 不相交, 而

$$P\{\eta<a\}=P\{f(\xi)<a\}=P\{\xi\in\sum_i E_i(a)\}=\sum_i\int_{E_i(a)}p(x)dx$$

$$=\sum_i\int_{-\infty}^{a}p[h_i(y)]|h_i'(y)|dy$$

$$=\int_{-\infty}^{a}\sum_i p[h_i(y)]|h_i'(y)|dy$$

因此 η 是具有密度函数(12)的连续型随机变量.

这结果包含了单调的场合(10), 此外, 例2也是其特殊场合. 再看下面例子.

［例 4］ 若 ξ 服从 $N(0,1)$, 计算 $\eta=\xi^2$ 的密度函数.

［解］ $y=f(x)=x^2$, 分段单调, 在 $(-\infty, 0)$ 中反函数 $x=h_1(y)=-\sqrt{y}$, 而在$[0, \infty)$中反函数为 $x=h_2(y)=\sqrt{y}$, 因此根

据(12)知道 η 的密度函数为

$$g(y)=\varphi(-\sqrt{y})\left|-\frac{1}{2\sqrt{y}}\right|+\varphi(\sqrt{y})\left|\frac{1}{2\sqrt{y}}\right|$$

$$=\frac{1}{\sqrt{2\pi}}y^{-\frac{1}{2}}e^{-\frac{y}{2}},\ y>0 \tag{13}$$

顺便指出，(13)是下列分布当 $n=1$ 时的特例。

[χ^2-分布]　具有密度函数

$$p(x)=\frac{1}{2^{n/2}\Gamma\left(\frac{n}{2}\right)}x^{\frac{n}{2}-1}e^{-\frac{x}{2}},\ x>0 \tag{14}$$

的分布称为具有自由度 n 的 χ^2 分布.

χ^2 分布在数理统计中有重要应用. 与(1.46)中的 Γ-分布比较，我们可以知道 χ^2 分布是它的特例，其中取 $\lambda=\frac{1}{2}, r=\frac{n}{2}$，即为 $G\left(\frac{1}{2},\frac{n}{2}\right)$.

这里，我们要再提到一个有趣而重要的事实.

[均匀分布的特殊地位]　若随机变量 ξ 的分布函数为 $F(x)$，因为 $F(x)$ 是非降函数，对任意 $0\leqslant y\leqslant 1$，可定义

$$F^{-1}(y)=\inf\{x:F(x)>y\} \tag{15}$$

作为 $F(x)$ 的反函数.

下面考察随机变量 $\theta=F(\xi)$ 的分布，这里 $F(x)$ 是连续函数. 对 $0\leqslant x\leqslant 1$,

$$P\{\theta<x\}=P\{F(\xi)<x\}$$

$$=P\{\xi<F^{-1}(x)\}=F(F^{-1}(x))=x \tag{16}$$

即 $\theta=F(\xi)$ 服从[0, 1]均匀分布. 这个结论在统计中起重要作用.

反之，若 θ 服从[0, 1]均匀分布，对任意分布函数 $F(x)$，令

$$\xi=F^{-1}(\theta) \tag{17}$$

$$P\{\xi<x\}=P\{F^{-1}(\theta)<x\}=P\{\theta<F(x)\}=F(x)$$

因此 ξ 是服从分布函数 $F(x)$的随机变量.

这样，只要我们能产生[0，1]中均匀分布的随机变量的子样(观察值)，那么我们也就能通过(17)产生分布函数为 $F(x)$的随机变量的子样，这结论在蒙特卡洛方法中具有基本的重要性. 通常的做法是利用数学或物理的方法产生[0，1]中均匀分布随机变量的子样(称为**均匀分布随机数**)，再利用变换(17)得到任意分布的随机数.

***随机变量的存在性定理** 利用上述结果，我们可以给下面定理一个构造性的证明.

定理 若 $F(x)$是左连续的单调不减函数，且 $F(-\infty)=0$, $F(+\infty)=1$，则存在一个概率空间$(\Omega, \mathscr{F}, P)$及其上的随机变量$\xi(\omega)$，使 $\xi(\omega)$的分布函数正好是 $F(x)$.

[证明] 取 $\Omega=[0,1]$，再取 $\mathscr{F}$ 为[0，1]中波雷尔点集全体，而P取为直线上的勒贝格测度(它是长度概念的推广，但对一切波雷尔点集都有定义). 定义 $\theta(\omega)=\omega$，则 $\theta(\omega)$是$(\Omega, \mathscr{F}, P)$上的随机变量，又对一切 $0\leqslant x\leqslant 1$,

$$P\{\theta(\omega)<x\}=P\{\omega\in[0,x)\}=x$$

因此 $\theta(\omega)$服从[0,1]上均匀分布.

再利用(15)定义 $F^{-1}(y)$，当然它也是单调函数，从而是波雷尔函数，令

$$\xi(\omega)=F^{-1}(\theta(\omega))$$

则 $\xi(\omega)$是$(\Omega, \mathscr{F}, P)$上的随机变量，而且由上段讨论可知，它的分布函数正好是 $F(x)$.

三、随机向量的函数的分布律

若 $\eta=f(\xi_1, \cdots, \xi_n)$，而$(\xi_1, \cdots, \xi_n)$的密度函数为$p(x_1, \cdots, x_n)$，则同上面一样讨论可以得到

$$G(y)=P\{\eta<y\}=\int\cdots\int_{f(x_1,\cdots,x_n)<y} p(x_1,\cdots,x_n)dx_1\cdots dx_n \quad (18)$$

我们看一些例子.

[和的分布] 若$\eta=\xi_1+\xi_2$，而(ξ_1,ξ_2)的密度函数为$p(x_1,x_2)$，则

$$F(y)=P\{\eta<y\}=\iint_{x_1+x_2<y} p(x_1,x_2)dx_1dx_2$$

$$=\int_{-\infty}^{\infty}\int_{-\infty}^{y-x_1} p(x_1,x_2)dx_1dx_2 \quad (19)$$

特别当ξ_1，ξ_2相互独立时，有$p(x_1,x_2)=p_1(x_1)p_2(x_2)$，这里$p_1(x_1)$为$\xi_1$的密度函数，$p_2(x_2)$为$\xi_2$的密度函数．代入(19)得

$$F(y)=\int_{-\infty}^{\infty}dx_1\int_{-\infty}^{y-x_1}p_1(x_1)p_2(x_2)dx_2$$

$$=\int_{-\infty}^{\infty}dx_1\int_{-\infty}^{y}p_1(x_1)p_2(z-x_1)dz$$

$$=\int_{-\infty}^{y}\left[\int_{-\infty}^{\infty}p_1(x_1)p_2(z-x_1)dx_1\right]dz$$

因此η的密度函数为

$$p(x)=\int_{-\infty}^{\infty}p_1(u)p_2(x-u)du \quad (20)$$

也可写为

$$p(x)=\int_{-\infty}^{\infty}p_1(x-u)p_2(u)du \quad (21)$$

(20)或(21)称为褶积公式(或卷积公式)，在概率论中相当重要.

[商的分布] 若$\eta=\dfrac{\xi_1}{\xi_2}$，而$(\xi_1,\xi_2)$的密度函数为$p(x_1,x_2)$，则

$$G(x)=P\{\eta<x\}=P\left\{\frac{\xi_1}{\xi_2}<x\right\}=\iint_{x_1/x_2<x}p(x_1,x_2)dx_1dx_2$$

$$=\int_0^{\infty}\int_{-\infty}^{zx} p(y,z)dydz+\int_{-\infty}^{0}\int_{zx}^{\infty} p(y,z)dydz \quad (22)$$

η 的密度函数为

$$g(x)=\int_0^{\infty} p(zx,z)zdz-\int_{-\infty}^{0} p(zx,z)zdz$$

$$=\int_{-\infty}^{\infty} |z|\, p(zx,z)dz \quad (23)$$

*[关于顺序统计量的若干分布]　若 $\xi_1, \xi_2, \cdots, \xi_n$ 是相互独立的随机变量，具有相同的分布，这里的 ξ_n^* 及 ξ_1^* 相当于把 ξ_1，ξ_2，$\cdots, \xi_n$ 按大小顺序重新排列为

$$\xi_1^* \leqslant \xi_2^* \leqslant \cdots \leqslant \xi_n^* \quad (24)$$

的末项及首项，它们在非参数统计中有重要应用。下面讨论几种与它们有关的分布。

首先求 ξ_n^* 的分布函数。

$$\begin{aligned} P\{\xi_n^*<x\} &= P\{\max(\xi_1, \xi_2, \cdots, \xi_n)<x\} \\ &= P\{\xi_1<x, \xi_2<x, \cdots, \xi_n<x\} \\ &= P\{\xi_1<x\}\cdot P\{\xi_2<x\}\cdots P\{\xi_n<x\} \\ &= [F(x)]^n \end{aligned} \quad (25)$$

其次求 ξ_1^* 的分布函数。注意到

$$\begin{aligned} P\{\xi_1^*\geqslant x\} &= P\{\min(\xi_1, \xi_2, \cdots, \xi_n)\geqslant x\} \\ &= P\{\xi_1\geqslant x, \xi_2\geqslant x, \cdots, \xi_n\geqslant x\} \\ &= P\{\xi_1\geqslant x\}P\{\xi_2\geqslant x\}\cdots P\{\xi_n\geqslant x\} \\ &= [1-F(x)]^n \end{aligned}$$

因此

$$P\{\xi_1^*<x\}=1-[1-F(x)]^n \quad (26)$$

进一步，讨论(ξ_1^*, ξ_n^*)的联合分布。

记 $G(x,y)=P\{\xi_1^*<x, \xi_n^*<y\}$。

若 $x\geqslant y$，则

$$G(x,y)=P\{\xi_1^*<x, \xi_n^*<y\}$$
$$=P\{\xi_n^*<y\}=[F(y)]^n \tag{27}$$

若 $x<y$，则

$$G(x,y)=P\{\xi_1^*<x, \xi_n^*<y\}$$
$$=P\{\xi_n^*<y\}-P\{\xi_1^*\geqslant x, \xi_n^*<y\}$$
$$=[F(y)]^n-[F(y)-F(x)]^n \tag{28}$$

其联合密度函数为

$$g(x,y)=\begin{cases}0, & x\geqslant y\\ n(n-1)[F(y)-F(x)]^{n-2}f(x)f(y), & x<y\end{cases} \tag{29}$$

最后，我们来求极差 $R=\xi_n^*-\xi_1^*$ 的分布密度函数 $f_R(r)$，显然对 $r\leqslant 0$，$f_R(r)=0$，若 $r>0$，则

$$P\{R<r\}=\iint_{y-x<r} g(x,y)dxdy$$
$$=\int_{-\infty}^{\infty}\left[\int_{-\infty}^{x+r} g(x,y)dy\right]dx$$
$$=\int_{-\infty}^{\infty}\left[\int_{-\infty}^{r} g(x,x+z)dz\right]dx$$
$$=\int_{-\infty}^{r}\left[\int_{-\infty}^{\infty} g(x,x+z)dx\right]dz$$

因此

$$f_R(r)=\int_{-\infty}^{\infty} g(x,x+r)dx$$
$$=n(n-1)\int_{-\infty}^{\infty}[F(x+r)-F(x)]^{n-2}$$
$$\cdot f(x)f(x+r)dx \tag{30}$$

四、随机向量的变换

若 $(\xi_1,\cdots,\xi_n)$ 的密度函数为 $p(x_1,\cdots,x_n)$，求 $\eta_1=f_1(\xi_1,\cdots,$

$\xi_n), \cdots, \eta_m = f_m(\xi_1, \cdots, \xi_n)$的分布．这时有

$$G(y_1, \cdots, y_m) = P\{\eta_1 < y_1, \cdots, \eta_m < y_m\}$$

$$= \underset{\substack{f_1(x_1, \cdots, x_n) < y_1 \\ \cdots \\ f_m(x_1, \cdots, x_n) < y_m}}{\int \cdots \int} p(x_1, \cdots, x_n) dx_1 \cdots dx_n \qquad (31)$$

显然，这是最一般的场合．当 $m=1$ 时便是随机向量的函数的情形，当 $m=n=1$ 时得到单个随机变量的函数的情形．下面考虑另一个重要的特殊情形，即当$(\xi_1, \cdots, \xi_n)$与$(\eta_1, \cdots, \eta_m)$有一一对应变换关系时，当然这时 $n=m$ 必须成立．

若对 $y_i = f_i(x_1, \cdots, x_n)$存在唯一的反函数 $x_i(y_1, \cdots, y_n) = x_i$，$(i=1, \cdots, n)$，且$(\eta_1, \cdots, \eta_n)$的密度函数为 $q(y_1, \cdots, y_n)$，那么

$$G(y_1, \cdots, y_n) = \underset{\substack{u_1 < y_1 \\ \cdots \\ u_n < y_n}}{\int \cdots \int} q(u_1, \cdots, u_n) du_1 \cdots du_n \qquad (32)$$

比较(31)与(32)可知

$$q(y_1, \cdots, y_n)$$
$$= \begin{cases} p(x_1, \cdots, x_n)|J|, & \text{若}(y_1, \cdots, y_n)\text{属于} f_1, \cdots, f_n \text{的值域} \\ 0, & \text{其他} \end{cases} \qquad (33)$$

其中 J 为坐标变换的雅可比行列式

$$J = \begin{vmatrix} \dfrac{\partial x_1}{\partial y_1} & \cdots & \dfrac{\partial x_n}{\partial y_1} \\ \cdots & \cdots & \cdots \\ \dfrac{\partial x_1}{\partial y_n} & \cdots & \dfrac{\partial x_n}{\partial y_n} \end{vmatrix} \qquad (34)$$

这里，我们假定上述偏导数存在而且连续．

公式(33)对应于单变量场合的公式(10)，也可以导出对应于公式(12)的多变量场合的公式，这留给读者作为练习。

〔例5〕 若(ξ_1, ξ_2)的密度函数为$p(x_1, x_2)$，而

$$\begin{aligned}\eta_1 &= a\xi_1 + b\xi_2 \\ \eta_2 &= c\xi_1 + d\xi_2\end{aligned} \tag{35}$$

这里 $\Delta = \begin{vmatrix} a & b \\ c & d \end{vmatrix} \neq 0$，试求$(\eta_1, \eta_2)$的密度函数$q(y_1, y_2)$。

〔解〕 在这个例中

$$y_1 = f_1(x_1, x_2) = ax_1 + bx_2, \qquad y_2 = f_2(x_1, x_2) = cx_1 + dx_2$$

因此

$$x_1 = \frac{d}{\Delta}y_1 - \frac{b}{\Delta}y_2, \quad x_2 = -\frac{c}{\Delta}y_1 + \frac{a}{\Delta}y_2$$

而

$$J = \begin{vmatrix} \dfrac{d}{\Delta} & -\dfrac{b}{\Delta} \\ -\dfrac{c}{\Delta} & \dfrac{a}{\Delta} \end{vmatrix} = \frac{ad-bc}{\Delta^2} = \frac{1}{ad-bc}$$

最后得到

$$q(y_1, y_2) = \frac{p\left(\dfrac{d}{\Delta}y_1 - \dfrac{b}{\Delta}y_2, -\dfrac{c}{\Delta}y_1 + \dfrac{a}{\Delta}y_2\right)}{|ad-bc|} \tag{36}$$

〔例6〕 若ξ，η为独立随机变量，且具有相同的指数分布密度函数

$$p(x) = \begin{cases} e^{-x}, & x \geqslant 0 \\ 0, & x < 0 \end{cases}$$

试求$\alpha = \xi + \eta$与$\beta = \dfrac{\xi}{\eta}$的密度函数$q(u, v)$。

〔解〕 对$u \geqslant 0, v \geqslant 0$作变换$u = x + y$，$v = \dfrac{x}{y}$，因此$x = \dfrac{uv}{1+v}$，$y = \dfrac{u}{1+v}$，

$$J^{-1}=\begin{vmatrix}\frac{\partial u}{\partial x} & \frac{\partial u}{\partial y}\\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y}\end{vmatrix}=\begin{vmatrix}1 & 1\\ \frac{1}{y} & -\frac{x}{y^2}\end{vmatrix}=\frac{-(x+y)}{y^2}=-\frac{(1+v)^2}{u}$$

所以

$$|J|=\frac{u}{(1+v)^2}, \qquad u\geqslant 0$$

$$q=p(x)p(y)|J|=e^{-(x+y)}|J|=e^{-u}\frac{u}{(1+v)^2}$$

$$=ue^{-u}\frac{1}{(1+v)^2} \tag{37}$$

意外的是，这里的 α 与 β 是独立的，分别具有密度 ue^{-u} 及 $\frac{1}{(1+v)^2}$.

五、随机变量的函数的独立性

首先证明一个定理.

定理 若 $\xi_1, \cdots, \xi_n$ 是相互独立的随机变量，则 $f_1(\xi_1), \cdots, f_n(\xi_n)$ 也是相互独立的，这里 $f_i(i=1,\cdots,n)$ 是任意的一元波雷尔函数.

[证明] 对任意的一维波雷尔点集 $B_1, \cdots, B_n$ 有

$$\begin{aligned}&P\{f_1(\xi_1)\in B_1, \cdots, f_n(\xi_n)\in B_n\}\\ &=P\{\xi_1\in f_1^{-1}(B_1), \cdots, \xi_n\in f_n^{-1}(B_n)\}\\ &=P\{\xi_1\in f_1^{-1}(B_1)\}\cdots P\{\xi_n\in f_n^{-1}(B_n)\}\\ &=P\{f_1(\xi_1)\in B_1\}\cdots P\{f_n(\xi_n)\in B_n\}\end{aligned}$$

定理的结论在直观上是明显的，但在定理的证明中却要两次用到未证明的论断(2.37)，其中第一次用来指明对 $\xi_1, \cdots, \xi_n$ 的有关概率可以化为乘积的形式，另一次用来说明最后的等式表明 $f_1(\xi_1), \cdots, f_n(\xi_n)$ 是相互独立的，且第一次是难以避免的.

例 6 说明，即使由相同的随机向量构成的不同函数也可能是独立的，这种情况在概率论与数理统计中相当重要，下面再讨论一

些例子.

[例 7] 若 ξ 与 η 是相互独立的随机变量，均服从 $N(0,1)$，试证 $\rho=\sqrt{\xi^2+\eta^2}$ 及 $\varphi=\mathrm{tg}^{-1}\left(\frac{\eta}{\xi}\right)$ 是相互独立的.

[解] 采用极坐标，$x=r\cos\theta$, $y=r\sin\theta$, 因此 $r=\sqrt{x^2+y^2}$, $\theta=\mathrm{tg}^{-1}\frac{y}{x}$, 因为 (ξ,η) 的密度函数为

$$p(x,y)=\frac{1}{2\pi}e^{-(x^2+y^2)/2}$$

而

$$J=\begin{vmatrix}\frac{\partial x}{\partial r} & \frac{\partial x}{\partial\theta}\\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial\theta}\end{vmatrix}=\begin{vmatrix}\cos\theta & -r\sin\theta\\ \sin\theta & r\cos\theta\end{vmatrix}=r$$

故 (ρ,φ) 的密度函数为

$$\begin{aligned}q(r,\theta)&=\frac{1}{2\pi}e^{-(x^2+y^2)/2}\cdot r\\&=\frac{1}{2\pi}\cdot re^{-r^2/2},\quad r\geqslant 0,0\leqslant\theta\leqslant 2\pi\end{aligned}\tag{38}$$

即 $\rho=\sqrt{\xi^2+\eta^2}$ 的密度函数为

$$R(r)=\begin{cases}re^{-r^2/2}, & r\geqslant 0\\ 0, & r<0\end{cases}\tag{39}$$

这个分布称为瑞利(Rayleigh)分布.

而 $\varphi=\mathrm{tg}^{-1}\frac{\eta}{\xi}$ 服从 $[0,2\pi]$ 中的均匀分布，而且它们是独立的.

这个结果常被用来产生服从正态分布的随机数. 做法如下：产生相互独立的[0,1]均匀分布的随机数 U_1, U_2, 令

$$\begin{aligned}\xi&=(-2\ln U_1)^{\frac{1}{2}}\cos 2\pi U_2\\ \eta&=(-2\ln U_1)^{\frac{1}{2}}\sin 2\pi U_2\end{aligned}\tag{40}$$

则 ξ 与 η 是相互独立的 $N(0,1)$ 随机数.

这样做法的理由让读者自行论证.

[例 8] 若 (ξ_1, ξ_2) 服从二元正态分布 (2.22), 其中 $a=0$, $b=0$. 令

$$\eta_1=\xi_1\cos\alpha+\xi_2\sin\alpha, \quad \eta_2=-\xi_1\sin\alpha+\xi_2\cos\alpha \tag{41}$$

这里 $0\leqslant\alpha\leqslant2\pi$, 是某个角度. 我们来求 (η_1, η_2) 的密度函数 $q(u, v)$.

这里可直接用例 5 的结果. 其中 $J=1$, 因此

$$q(u, v)=p(u\cos\alpha-v\sin\alpha, u\sin\alpha+v\cos\alpha)$$

$$=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}}\exp\left\{-\frac{1}{2(1-r^2)}(Au^2-2Buv+Cv^2)\right\} \tag{42}$$

其中

$$A=\frac{\cos^2\alpha}{\sigma_1^2}-2r\frac{\cos\alpha\sin\alpha}{\sigma_1\sigma_2}+\frac{\sin^2\alpha}{\sigma_2^2}$$

$$B=\frac{\cos\alpha\sin\alpha}{\sigma_1^2}-r\frac{\sin^2\alpha-\cos^2\alpha}{\sigma_1\sigma_2}-\frac{\cos\alpha\sin\alpha}{\sigma_2^2}$$

$$C=\frac{\sin^2\alpha}{\sigma_1^2}+2r\frac{\cos\alpha\sin\alpha}{\sigma_1\sigma_2}+\frac{\cos^2\alpha}{\sigma_2^2}$$

由(42)可看出由二维正态变量 (ξ_1, ξ_2) 经坐标旋转而得的随机向量 (η_1, η_2) 还是服从正态分布. 进一步, 若选 α 使得

$$\mathrm{tg}\,2\alpha=\frac{2r\sigma_1\sigma_2}{\sigma_1^2-\sigma_2^2} \tag{43}$$

则 $B=0$, 因此 η_1 与 η_2 独立. 这说明二元正态分布可经适当的坐标旋转化为两个独立正态分布之积.

第三章小结

本章中我们详细地研究了随机变量. 用随机变量描述随机现象是近代概率论中最重要的方法, 以后我们所讨论的随机事件几

乎没有例外的都用随机变量来描述.

我们给出了随机变量的严格定义，按照这种观点，随机变量是定义在样本空间上的具有某种可测性的实值函数，只是出于历史的原因，才沿用"变量"二字.

对于随机变量，重要的是要知道它取哪一些值以及以怎样的概率取这些值. 从这个角度讲，分布函数完整地描述了随机变量. 同时，分布函数具有良好的分析性质，很便于研究，它是研究随机变量的重要工具.

离散型随机变量与连续型随机变量是最重要的两类随机变量，由于它们的取值范围不同，因此对它们的描述及处理方法都有很大不同，应该进行对比，从而加深理解.

我们还介绍了多种分布函数，而且比较注重各种分布函数的特征性质. 这些分布函数在不同的理论和实际问题中扮演着重要角色.

正态分布是概率论中最重要的分布，在应用中及理论研究中占有头等重要的地位，它与二项分布及普阿松分布是概率论中最重要的三种分布. 判断一种分布重要性的标准是：(1) 在实际工作中经常碰到；(2)在理论研究中重要，具有较好的性质；(3)用它能导出许多重要分布. 以上三种分布都满足这些要求.

把几个随机变量放在一起研究时不但需要研究各个分量个别的性质，而且要考虑它们之间的联系，从而大大丰富了研究的内容，条件分布及独立性概念也随之出现，它们是条件概率及事件独立性概念在随机变量场合的具体化，在今后研究中很重要.

随机变量的函数的分布律的推导，在数理统计中及在概率论的许多应用中相当重要. 我们这里分三种类型介绍了最基本的处理方法. 这部分内容，只有通过动手多作练习，才能牢固掌握.

习　题

1. 直线上有一质点，每经一个单位时间，它分别以概率 p 及 $1-p$ 向右或向左移动一格，若该质点在时刻 0 从原点出发，而且每次移动是相互独立，试用随机变量来描述这质点的运动(以 S_n 表示时刻 n 时质点的位置)。

2. 设 ξ 为贝努里试验中第一个游程(连续的成功或失败)的长，试求 ξ 的概率分布。

3. C 应取何值才能使下列函数成为概率分布：

(1) $f(k)=\dfrac{C}{N}, k=1, 2, \cdots, N$； (2) $f(k)=C\dfrac{\lambda^k}{k!}, k=1, 2, \cdots, \lambda>0$。

4. 证明函数

$$f(x)=\frac{1}{2}e^{-|x|}, \ -\infty<x<\infty$$

是一个密度函数。

5. 若 ξ 的分布函数为 $N(10, 4)$，求 ξ 落在下列范围的概率。(1)$(6, 9)$；(2)$(7, 12)$；(3)$(13, 15)$。

6. 若 ξ 的分布函数为 $N(60, 9)$，求分点 x_1, x_2, x_3, x_4 使 ξ 落在 $(-\infty, x_1)$, (x_1, x_2), (x_2, x_3), (x_3, x_4), (x_4, ∞) 中的概率之比为 7:24:38:24:7.

7. 若 ξ 的分布函数为 $N(5, 4)$，求 a 使：(1)$P\{\xi<a\}=0.90$；(2)$P\{|\xi-5|>a\}=0.01$.

8. 设 $F(x)=P\{\xi\leqslant x\}$，试证 $F(x)$ 具有下列性质：(1) 非降；(2) 右连续；(3)$F(-\infty)=0, F(+\infty)=1$.

9. 试证：若 $P\{\xi\leqslant x_2\}\geqslant 1-\beta$, $P\{\xi\geqslant x_1\}\geqslant 1-\alpha$，则 $P\{x_1\leqslant\xi\leqslant x_2\}\geqslant 1-(\alpha+\beta)$.

10. 设随机变量 ξ 取值于$[0, 1]$，若 $P\{x\leqslant\xi<y\}$ 只与长度 $y-x$ 有关(对一切 $0\leqslant x\leqslant y\leqslant 1$)，试证 ξ 服从$[0, 1]$均匀分布。

*11. 若存在 Θ 上的实值函数 $Q(\theta)$ 及 $D(\theta)$ 以及 $T(x)$ 及 $S(x)$，使

$$f_\theta(x)=\exp\{Q(\theta)T(x)+D(\theta)+S(x)\}$$

则称 $\{f_\theta, \theta\in\Theta\}$ 是一个单参数的指数族。证明(1)正态分布 $N(m_0, \sigma^2)$，已知 m_0，关于参数 σ；(2)正态分布 $N(m, \sigma_0^2)$，已知 σ_0，关于参数 m；(3)普阿松分布 $p(k;\lambda)$ 关于 λ 是一个单参数的指数族。

但是$[0, \theta]$上均匀分布，关于 θ 不是一个单参数的指数族。

12. 定义二元函数

$$F(x, y)=\begin{cases}1, & x+y>0 \\ 0, & x+y\leq 0\end{cases}$$

验证此函数对每个变元非降，左连续，且满足(2.6)及(2.7)，但无法使(2.5)保持非负。

13. 试证 $f(x, y)=Ke^{-(ax^2+2bxy+cy^2)}$ 为密度函数的充要条件为 $a>0$, $c>0$, $b^2-ac<0$, $K=\frac{1}{\pi}\sqrt{ac-b^2}$.

14. 若$f_1(x)$, $f_2(y)$为分布密度，求为使$f(x, y)=f_1(x)\times f_2(y)+h(x, y)$成为密度函数，$h(x, y)$必须而且只须满足什么条件。

15. 若(ξ, η)的密度函数为

$$f(x, y)=\begin{cases}Ae^{-(2x+y)}, & x>0, \quad y>0 \\ 0, & 其他\end{cases}$$

试求：(1)常数 A；(2)$P\{\xi<2, \eta<1\}$；(3)ξ 的边际分布；(4)$P\{\xi+\eta<$ (5)$f(x|y)$；(6)$P\{\xi<2|\eta<1\}$.

16. 若(ξ, η)服从二元正态分布，参数 $a, b, \sigma_1^2, \sigma_2^2, r$，以 $D(\lambda)$记下面圆的内部

$$\frac{(x-a)^2}{\sigma_1^2}-\frac{2r(x-a)(y-b)}{\sigma_1\sigma_2}+\frac{(y-b)^2}{\sigma_2^2}=\lambda^2$$

试求 $P\{(\xi, \eta)\in D(\lambda)\}$.

17. 证明多项分布的边际分布仍为多项分布。

18. 设二维随机变量(ξ, η)的联合密度为

$$p(x, y)=\frac{1}{\Gamma(k_1)\Gamma(k_2)}x^{k_1-1}(y-x)^{k_2-1}e^{-y}$$

$k_1>0, k_2>0, 0<x\leq y<\infty$. 试求 ξ 与 η 的边际分布。

19. 若 $f_1(x), f_2(x), f_3(x)$是对应于分布函数 $F_1(x), F_2(x), F_3(x)$的密度函数，证明对于一切 $\alpha(-1<\alpha<1)$，下列函数是密度函数，且具有相同的边际密度函数 $f_1(x), f_2(x), f_3(x)$：

$$f_\alpha(x_1, x_2, x_3)$$
$$=f_1(x_1)f_2(x_2)f_3(x_3)\{1+\alpha[2F_1(x_1)-1]\cdot[2F_2(x_2)-1][2F_3(x_3)-$$

20. 设 ξ 与 η 是相互独立的随机变量，均服从几何分布

$$g(k, p)=q^{k-1}p, \quad k=1, 2, \cdots$$

令$\zeta=\max(\xi,\eta)$，试求(1)(ζ,ξ)的联合分布；(2)ζ的分布；(3)ξ关于ζ的条件分布。

21. (1) 若(ξ,η)的联合密度函数为

$$f(x,y)=\begin{cases}4xy, & 0\leqslant x\leqslant 1,\quad 0\leqslant y\leqslant 1\\ 0, & 其他\end{cases}$$ 问ξ与η是否相互独立？

(2) 若(ξ,η)的联合密度函数为 $g(x,y)=\begin{cases}8xy, & 0\leqslant x\leqslant y,\ 0\leqslant y\leqslant 1\\ 0, & 其他\end{cases}$ 问ξ与η是否相互独立？

22. 设(ξ,η,ζ)的联合密度函数为

$$p(x,y,z)=\begin{cases}\dfrac{1}{8\pi^3}(1-\sin x\sin y\sin z), & 当\ \begin{array}{l}0\leqslant x\leqslant 2\pi\\ 0\leqslant y\leqslant 2\pi\\ 0\leqslant z\leqslant 2\pi\end{array}\ 时\\ 0, & 其他\end{cases}$$

试证ξ,η,ζ两两独立，但不相互独立。

23. 设(ξ,η)具有联合密度函数

$$p(x,y)=\begin{cases}\dfrac{1+xy}{4}, & |x|<1,\quad |y|<1\\ 0, & 其他\end{cases}$$

试证ξ与η不独立，但ξ^2与η^2是相互独立的。

24. 若ξ_1与ξ_2是独立随机变量，均服从普阿松分布，参数为λ_1及λ_2，试直接证明

(1) $\xi_1+\xi_2$具有普阿松分布，参数为$\lambda_1+\lambda_2$；

(2) $P\{\xi_1=k|\xi_1+\xi_2=n\}=\binom{n}{k}\left(\dfrac{\lambda_1}{\lambda_1+\lambda_2}\right)^k\left(\dfrac{\lambda_2}{\lambda_1+\lambda_2}\right)^{n-k}$.

25. 若ξ,η相互独立且皆以概率$\frac{1}{2}$取值$+1$及-1，令$\zeta=\xi\eta$，试证ξ,η,ζ两两独立但不相互独立。

26. 若ξ服从普阿松分布，参数为λ，试求(1)$\eta=a\xi+b$及(2)$\eta=\xi^2$的分布。

27. 设ξ的密度函数为$p(x)$，求下列随机变量的分布函数：(1)$\eta=\xi^{-1}$，这里$P\{\xi=0\}=0$；(2)$\eta=\operatorname{tg}\xi$；(3)$\eta=|\xi|$。

28. 对圆的直径作近似度量，设其值均匀分布于(a,b)内，试求圆面积的

分布密度。

29. 若ξ，η为相互独立的分别服从[0.1]均匀分布的随机变量，试求$\zeta=\xi+\eta$的分布密度函数。

30. 设ξ,η相互独立，分别服从$N(0,1)$，试求$\zeta=\frac{\xi}{\eta}$的密度函数。

31. 若ξ,η是独立随机变量，均服从$N(0,1)$，试求$U=\xi+\eta$，$V=\xi-\eta$的联合密度函数。

32. 若ξ_1，ξ_2，…，ξ_n相互独立，且皆服从指数分布，参数分别为λ_1，$\lambda_2,\cdots,\lambda_n$，试求$\eta=\min(\xi_1,\xi_2,\cdots,\xi_n)$的分布。

33. 在$(0,a)$线段上随机投掷两点，试求两点间距离的分布函数。

34. 若气体分子的速度是随机向量$V=(X,Y,Z)$，各分量相互独立，且均服从$N(0,\sigma^2)$，试证$S=\sqrt{X^2+Y^2+Z^2}$服从**马克斯威尔分布**

$$f(s)=\sqrt{\frac{2}{\pi}}\ \frac{s^2}{\sigma^3}\exp\left(-\frac{s^2}{2\sigma^2}\right),\quad s>0$$

35. 设ξ,η为两个独立随机变量，ξ服从$N(0,1)$，η服从自由度为n的χ^2分布(3.14)，令$t=\xi\Big/\sqrt{\frac{\eta}{n}}$，试证$t$的密度函数为

$$p_n(x)=\frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)}\left(\frac{x^2}{n}+1\right)^{-(n+1)/2}$$

这分布称为**具有自由度n的t-分布**，在数理统计中十分重要。

36. 设ξ,η,ζ有联合密度函数

$$f(x,y,z)=\begin{cases}6(1+x+y+z)^{-4}, & \text{当 } x>0,y>0,z>0 \text{ 时}\\ 0, & \text{其他}\end{cases}$$

试求$U=\xi+\eta+\zeta$的密度函数。

37. 若ξ，η独立，且均服从$N(0,1)$，试证$U=\xi^2+\eta^2$与$V=\frac{\xi}{\eta}$是独立的。

38. 求证，如果ξ与η独立，且分别服从Γ-分布$G(\lambda,r_1)$和$G(\lambda,r_2)$，则$\xi+\eta$与$\frac{\xi}{\eta}$也独立。

39. 设独立随机变量ξ,η均服从

$$p(x)=\begin{cases}e^{-x}, & x>0\\ 0, & 其他\end{cases}$$

问 $\xi+\eta$ 与 $\dfrac{\xi}{\xi+\eta}$ 是否独立？

40. 若 (ξ,η) 服从二元正态分布(2.22)，试找出 $\xi+\eta$ 与 $\xi-\eta$ 相互独立的充要条件。

41. 对二元正态密度函数

$$p(x,y)=\frac{1}{2\pi}\exp\left\{-\frac{1}{2}(2x^2+y^2+2xy-22x-14y+65)\right\}$$

(1)把它化为标准形式(2.22)；(2)指出 a,b,σ_1,σ_2,r；(3)求 $p_1(x)$；(4)求 $p(x|y)$。

42. 设 $a=0$，$B^{-1}=\begin{pmatrix}7&3&2\\3&4&1\\2&1&2\end{pmatrix}$，试写出分布密度(2.12)，并求出 (ξ_1,ξ_2) 的边际密度函数。

*43. 设 ξ 与 η 是相互独立相同分布的随机变量，其密度函数不等于 0 且有二阶导数，试证若 $\xi+\eta$ 与 $\xi-\eta$ 相互独立，则随机变量 $\xi,\eta,\xi+\eta,\xi-\eta$ 均服从正态分布。

*44. 利用随机变量分布解释贝特朗奇论。

*45. 若 f 是 Ω 上单值实函数，对 $B\subset R^1$，记 $f^{-1}(B)=\{\omega\in\Omega:f(\omega)\in B\}$，试证逆映照 f^{-1} 具有如下性质：

(1) $f^{-1}\left(\bigcup\limits_{\lambda\in\Lambda}B_\lambda\right)=\bigcup\limits_{\lambda\in\Lambda}f^{-1}(B_\lambda)$；

(2) $f^{-1}\left(\bigcap\limits_{\lambda\in\Lambda}B_\lambda\right)=\bigcap\limits_{\lambda\in\Lambda}f^{-1}(B_\lambda)$；

(3) $f^{-1}(\overline{B})=\overline{f^{-1}(B)}$。

*46. 证明：ξ 是一个随机变量当且仅当对任何 $x\in R^1$，成立

$$\{\omega:\xi(\omega)<x\}\in\mathscr{F}$$

[提示：必要性是显然的，为证充分性，记 $\mathfrak{M}=\{A:A\subset R^1,\ (\xi(\omega)\in A)\in\mathscr{F}\}$，验证 $\mathfrak{M}$ 是 σ-域，又 $\mathfrak{M}$ 包含全体形如 $(-\infty,\ x)$ 的区间，故 $\mathfrak{M}$ 包含 $\mathscr{B}$]

第四章　数字特征与特征函数

§1. 数学期望，方差，矩

一、数学期望

有甲、乙两个射手，他们的射击技术用下表表出：

甲射手

击中环数	8	9	10
概　　率	0.3	0.1	0.6

乙射手

击中环数	8	9	10
概　　率	0.2	0.5	0.3

试问哪一个射手本领较好？

这个问题的答案不是一眼看得出的．这说明了分布列虽然完整地描述了随机变量，但是却不够"集中"地反映出它的变化情况．因此我们有必要找出一些量来更集中、更概括地描述随机变量，这些量多是某种平均值．

在上面问题中，若使两个射手各射N枪，则他们打中的环数大约是：

甲：　$8\times0.3N+9\times0.1N+10\times0.6N=9.3N$；

乙：　$8\times0.2N+9\times0.5N+10\times0.3N=9.1N$．

平均起来甲每枪射中 9.3 环，乙射中 9.1 环，因此甲射手的本领要好些．

受上面问题启发，对一般离散型随机变量，我们可引进如下定义．

定义　设ξ为一离散型随机变量，它取值$x_1, x_2, x_3, \cdots$，对应的概率为$p_1, p_2, p_3, \cdots$，如果级数

$$\sum_{i=1}^{\infty} x_i p_i \tag{1}$$

绝对收敛，则把它称为 ξ 的数学期望或均值，记作 $E\xi$。

当 $\sum_{i=1}^{\infty}|x_i|p_i$ 发散时，则说 ξ 的数学期望不存在。

定义中对级数要求绝对收敛是为了数学处理的方便．从直观上来讲，它也是合理的：因为诸 x_i 的顺序对随机变量并不是本质的，因而在数学期望的定义中就应允许任意改变 x_i 的次序而不影响其收敛性及其和值，这在数学上就相当于要求级数（1）绝对收敛。

显然数学期望由概率分布唯一确定，以后我们也称为某概率分布的数学期望．下面来计算一些重要的离散型分布的期望值。

［例 1］ 二项分布 $p_k=\binom{n}{k}p^k q^{n-k}$, $k=0,1,2,\cdots,n$

$$\sum_{k=0}^{n} k p_k=\sum_{k=1}^{n} k\binom{n}{k}p^k q^{n-k}$$

$$=np\sum_{k=1}^{n}\binom{n-1}{k-1}p^{k-1}q^{n-k}$$

$$=np(p+q)^{n-1}=np \tag{2}$$

［例 2］ 普阿松分布 $p_k=\frac{\lambda^k}{k!}e^{-\lambda}$, $k=0,1,2,\cdots$

$$\sum_{k=0}^{\infty} k p_k=\sum_{k=1}^{\infty} k\cdot\frac{\lambda^k}{k!}e^{-\lambda}=\lambda e^{-\lambda}\sum_{k=1}^{\infty}\frac{\lambda^{k-1}}{(k-1)!}$$

$$=\lambda e^{-\lambda}\cdot e^{\lambda}=\lambda \tag{3}$$

由此看出，普阿松分布的参数 λ 就是它的期望值。

［例 3］ 几何分布 $\quad p_k=q^{k-1}p$, $\quad k=1,2,\cdots$

$$\sum_{k=1}^{\infty} kp_k=\sum_{k=1}^{\infty} kq^{k-1}p=p(1+2q+3q^2+\cdots)$$

$$=p(q+q^2+q^3+\cdots)'=p\left(\frac{q}{1-q}\right)'$$

$$=p\frac{1}{(1-q)^2}=\frac{1}{p} \tag{4}$$

[例 4] 随机变量 ξ 取值 $x_k=(-1)^k\frac{2^k}{k}, k=1,2,\cdots$对应的概率为 $p_k=\frac{1}{2^k}$，则由于 $p_k\geqslant0$; $\sum\limits_{k=1}^{\infty}p_k=1$，因此它是概率分布，而且

$$\sum_{k=1}^{\infty} x_k p_k=\sum_{k=1}^{\infty}(-1)^k\frac{1}{k}=-\ln 2$$

但由于

$$\sum_{k=1}^{\infty} |x_k| p_k=\sum_{k=1}^{\infty}\frac{1}{k}=\infty$$

因此 ξ 的数学期望不存在。

从上面例子看出，几种重要的离散型分布，其参数都可由数学期望算得，因此数学期望是一个重要的概念。下面例子说明数学期望概念是非常有用的。

[例 5] （一种验血新技术）在一个人数很多的单位中普查某种疾病，N个人去验血。对这些人的血的化验可以用两种办法进行。(1)每个人的血分别化验，这时需要化验N次；(2)把 k 个人的血混在一起进行化验。如果结果是阴性的，那么对这 k 个人只作一次检验就够了；如果结果是阳性的，那么必须对这 k 个人再逐个分别化验，这时对这 k 个人共需作 $k+1$ 次化验。假定对所有的人来说，化验是阳性反应的概率都是 p，而且这些人的反应是独立

的．我们来说明用办法(2)能减少化验的次数．

若记 $q=1-p$，则 k 个人的混血呈阳性反应的概率为 $1-q^k$，用(2)的方法验血时，每个人的血需要化验的次数 ξ 是随机变量，其分布列为

$$\begin{pmatrix} \frac{1}{k}, & 1+\frac{1}{k} \\ q^k, & 1-q^k \end{pmatrix}$$

因此

$$E\xi=\frac{1}{k}\cdot q^k+\left(1+\frac{1}{k}\right)(1-q^k)=1-q^k+\frac{1}{k}$$

N个人需要的化验次数的期望值为$N\left(1-q^k+\frac{1}{k}\right)$，当 $q^k-\frac{1}{k}>0$ 时，就能减少验血次数．例如当 $p=0.1$时，取$k=4$，则$q^k-\frac{1}{k}=0.4$．用(2)法平均能减少 40%的工作量．显然 p 愈小用这种方法愈有利．当 p 已知时，还可以选定整数 k_0，使 $E\xi$ 达到最小，把 k_0 个人分为一组就最能节省化验次数．

下面我们转入考虑连续型随机变量的数学期望．设随机变量 ξ 有密度函数 $p(x)$，取很密的分点 $x_0<x_1<x_2<\cdots<x_n$，则 ξ 落在 $[x_i, x_{i+1})$中的概率近似地等于 $p(x_i)(x_{i+1}-x_i)$，因此 ξ 与以概率 $p(x_i)(x_{i+1}-x_i)$取值 x_i 的离散型随机变量近似，而这离散型随机变量的数学期望为

$$\sum_i x_i p(x_i)(x_{i+1}-x_i)$$

上式是积分$\int_{-\infty}^{\infty} xp(x)dx$ 的渐近和式，这个直观的考虑启发我们引进如下定义：

定义 设 ξ 为具有密度函数 $p(x)$的连续型随机变量，当积分 $\int_{-\infty}^{\infty} xp(x)dx$ 绝对收敛时，我们称它为 ξ 的**数学期望(或均值)**，记

作 $E\xi$，即

$$E\xi=\int_{-\infty}^{\infty}xp(x)dx \tag{5}$$

显然这里定义的数学期望也只与分布有关．下面计算一些重要的连续型分布的数学期望．

［例 6］ 正态分布 $N(a, \sigma^2)$．

$$\begin{aligned}\int_{-\infty}^{\infty}xp(x)dx&=\int_{-\infty}^{\infty}x\frac{1}{\sqrt{2\pi}\,\sigma}e^{-(x-a)^2/2\sigma^2}dx\\&=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}(\sigma z+a)e^{-z^2/2}dz\\&=\frac{a}{\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{-z^2/2}dz=a\end{aligned} \tag{6}$$

可见 $N(a, \sigma^2)$ 中的参数 a 正是它的数学期望．

［例 7］ 指数分布 $p(x)=\lambda e^{-\lambda x}$，$x\geqslant 0$．

$$\int_0^{\infty}x\lambda e^{-\lambda x}dx=-\int_0^{\infty}xde^{-\lambda x}=\int_0^{\infty}e^{-\lambda x}dx=\frac{1}{\lambda} \tag{7}$$

［例 8］ 柯西分布 $p(x)=\frac{1}{\pi}\cdot\frac{1}{1+x^2}$．

由于

$$\int_{-\infty}^{\infty}|x|\cdot\frac{1}{\pi(1+x^2)}dx=\infty$$

因此柯西分布的数学期望不存在．

我们已经对离散型随机变量及连续型随机变量分别定义了数学期望，现在自然希望找到一种能适合一切随机变量的数学期望定义，并把上述两种情况作为特例．为了做到这点，需要利用斯蒂阶(stieltjes)积分①．

---

① 如果读者不熟悉斯蒂阶积分而又希望对它有所了解，可参看复旦大学数学系，《实变数函数论与泛函分析概要》(第二版)，上海科技出版社，1963，第四章．

不过，为读懂本书其余部分，只要承认后面所述的少数事实就可以了．

若随机变量 ξ 的分布函数为 $F(x)$，类似于连续型随机变量的场合，作很密的分割 $x_0<x_1<x_2<\cdots<x_n$，则 ξ 落在 $[x_i, x_{i+1})$ 中的概率等于 $F(x_{i+1})-F(x_i)$，因此 ξ 与以概率 $F(x_{i+1})-F(x_i)$ 取值 x_i 的离散型随机变量近似，而后者的数学期望为

$$\sum_i x_i[F(x_{i+1})-F(x_i)]$$

注意到上式是斯蒂阶积分 $\int_{-\infty}^{\infty} x dF(x)$ 的渐近和式，这启发我们引进如下定义：

定义 若 ξ 的分布函数为 $F(x)$，则定义

$$E\xi=\int_{-\infty}^{\infty} x dF(x) \tag{8}$$

为 ξ 的**数学期望**. 这里我们还是要求上述积分绝对收敛，否则数学期望不存在.

关于斯蒂阶积分

$$I=\int_{-\infty}^{\infty} g(x) dF(x) \tag{9}$$

我们仅列举它的如下性质：

(i) 当 $F(x)$ 为跳跃函数，在 $x_i (i=1, 2, \cdots)$ 具有跃度 p_i 时，上面积分化为无穷级数

$$I=\sum_i g(x_i) p_i$$

(ii) 当 $F(x)$ 存在导数 $F'(x)=p(x)$ 时，积分 (9) 化为普通积分

$$I=\int_{-\infty}^{\infty} g(x) p(x) dx$$

(iii) 线性性

$$\int_{-\infty}^{\infty}[a g_1(x)+b g_2(x)] dF(x)$$

$$=a\int_{-\infty}^{\infty}g_1(x)dF(x)+b\int_{-\infty}^{\infty}g_2(x)dF(x) \tag{10}$$

(iv)
$$\int_{-\infty}^{\infty}g(x)d[aF_1(x)+bF_2(x)]$$
$$=a\int_{-\infty}^{\infty}g(x)dF_1(x)+b\int_{-\infty}^{\infty}g(x)dF_2(x) \tag{11}$$

(v)
$$\int_a^b g(x)dF(x)=\int_a^c g(x)dF(x)+\int_c^b g(x)dF(x) \quad (a\leqslant c\leqslant b) \tag{12}$$

(vi) 若$g(x)\geqslant 0$, $F(x)$单调不减, $b>a$, 则
$$\int_a^b g(x)dF(x)\geqslant 0 \tag{13}$$

从前两个性质, 我们知道定义式(8)的确能包含(1)及(5)作为特例.

下面讨论随机变量的函数 $\eta=g(\xi)$ 的数学期望的定义, 这里 $g(x)$是一元波雷尔函数, 从上章 §3 的讨论已知η是随机变量. 通过类似于引进(8)的推理, 似应定义 $g(\xi)$的数学期望为
$$Eg(\xi)=\int_{-\infty}^{\infty}g(x)dF_\xi(x) \tag{14}$$
这里当然还是要求这个积分绝对收敛.

但是, 另一方面, 因为η是随机变量, 也有分布函数, 记之为 $F_\eta(x)$, 则按一般随机变量数学期望的定义(8)又应有
$$E\eta=\int_{-\infty}^{\infty}ydF_\eta(y) \tag{15}$$
因此, 这两个积分应该相等. 事实上, 这两个积分的确相等, 我们把这个事实写成定理的形式.

定理 1 若 $g(x)$是一元波雷尔函数, 而$\eta=g(\xi)$, 则
$$\int_{-\infty}^{\infty}ydF_\eta(y)=\int_{-\infty}^{\infty}g(x)dF_\xi(x) \tag{16}$$
即这两个积分中, 若有一个存在, 则另一个也存在, 而且两者相等.

这个定理的证明要用到积分论，超出了本课程范围．我们只能列出结论，并指出它的重要性：一方面，它消除了随机变量的数学期望定义中所出现的表面矛盾；另一方面在计算随机变量函数的数学期望时也带来很大的方便，我们无须先计算η的分布函数$F_\eta(y)$再求其数学期望，而可以直接从ξ的分布函数$F_\xi(x)$出发利用(14)来计算．

最重要的特殊情况，当然是ξ具有密度函数$p(x)$的场合，此时

$$E\eta=Eg(\xi)=\int_{-\infty}^{\infty}g(x)p(x)dx \tag{17}$$

这个结论经常用到，下面是一个例子，这个例子还顺带说明可以利用随机变量的期望值来作出某种最优决策．

［例 9］　假定在国际市场上每年对我国某种出口商品的需求量是随机变量ξ（单位吨），它服从[2000, 4000]均匀分布．设每售出这种商品一吨，可为国家挣得外汇 3 万元，但假如销售不出而屯积于仓库，则每吨需浪费保养费 1 万元，问题是要确定应组织多少货源，才能使国家的收益最大．

若以y记预备某年出口的此种商品量（显然可以只考虑$2000\leqslant y\leqslant 4000$的情况），则收益（单位：万元）

$$\eta=H(\xi)=\begin{cases}3y, & \text{当 } \xi\geqslant y \text{ 时}\\ 3\xi-(y-\xi), & \text{当 } \xi<y \text{ 时}\end{cases}$$

为了求得$E\eta$，利用公式(17)

$$\begin{aligned}E\eta&=\int_{-\infty}^{\infty}H(x)f(x)dx=\frac{1}{2000}\int_{2000}^{4000}H(x)dx\\&=\frac{1}{2000}\int_{2000}^{y}(4x-y)dx+\frac{1}{2000}\int_{y}^{4000}3ydx\\&=\frac{1}{1000}[-y^2+7000y-4\times10^6]\end{aligned}$$

此式当 $y=3500$ 时达到最大，因此组织 3500 吨此种商品是最好的决策。

可以把上面结果推广到随机向量的场合，若 $(\xi_1,\cdots,\xi_n)$ 的分布函数为 $F(x_1,\cdots,x_n)$，而 $g(x_1,\cdots,x_n)$ 为 n 元波雷尔函数，则

$$Eg(\xi_1,\cdots,\xi_n)=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}g(x_1,\cdots,x_n)dF(x_1,\cdots,x_n) \tag{18}$$

特别地，

$$E\xi_1=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}x_1dF(x_1,\cdots,x_n)=\int_{-\infty}^{\infty}x_1dF_1(x_1)$$

其中 $F_1(x_1)$ 是 ξ_1 的分布函数。一般地，引进如下定义

定义 随机向量 $(\xi_1,\xi_2,\cdots,\xi_n)$ 的数学期望为 $(E\xi_1,E\xi_2,\cdots,E\xi_n)$，其中

$$E\xi_i=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}x_idF(x_1,\cdots,x_n)=\int_{-\infty}^{\infty}x_idF_i(x_i) \tag{19}$$

这里 $F_i(x_i)$ 是 ξ_i 的分布函数。

因此二元正态分布(3.2.22)中的参数 (a,b) 正是随机向量的数学期望。

数学期望的基本性质

性质 1 若 $a\leqslant\xi\leqslant b$，则 $a\leqslant E\xi\leqslant b$。特别地 $Ec=c$，这里 a,b,c 是常数。

性质 2 线性性：对任意常数 c_i，$i=1,\cdots,n$，及 b，有

$$E\left(\sum_{i=1}^{n}c_i\xi_i+b\right)=\sum_{i=1}^{n}c_iE\xi_i+b \tag{20}$$

这两个性质的证明是明显的。性质 2 对数学期望的计算很有好处。

[例 10] 求超几何分布 $p_m=\dfrac{\binom{M}{m}\binom{N-M}{n-m}}{\binom{N}{n}}$, $m=0,1,\cdots,n$

的数学期望。

当然可以用$\sum_{k=0}^{n} kp_k$直接求出，但也可用下面方法来计算。

想一个相应的不放回抽样，令

$$\xi_i=\begin{cases}1, & \text{第 } i \text{ 次抽得废品}\\ 0, & \text{第 } i \text{ 次抽得好品}\end{cases}$$

则$P\{\xi_i=1\}=\frac{M}{N}$，因此$E\xi_i=\frac{M}{N}$，而$\xi=\xi_1+\cdots+\xi_n$表示n次抽样中抽出的废品数，它服从超几何分布，利用性质2得到

$$E\xi=E\xi_1+\cdots+E\xi_n=\frac{nM}{N} \tag{21}$$

二、方差

数学期望是随机变量的重要数字特征，它表示了随机变量的平均值．另一个重要的数字特征是方差．

定义 若$E(\xi-E\xi)^2$存在，则称它为随机变量ξ的方差，并记为$D\xi$，而称$\sqrt{D\xi}$为ξ的根方差或标准差.

利用(16)及(10)，

$$\begin{aligned}D\xi&=\int_{-\infty}^{\infty}(x-E\xi)^2dF_\xi(x)=\int_{-\infty}^{\infty}[x^2-2xE\xi+(E\xi)^2]dF_\xi(x)\\&=\int_{-\infty}^{\infty}x^2dF_\xi(x)-2E\xi\int_{-\infty}^{\infty}xdF_\xi(x)+(E\xi)^2=E\xi^2-(E\xi)^2\end{aligned} \tag{22}$$

方差描述了随机变量对于其数学期望的离散程度，在概率论和数理统计中十分重要．例如本节开头讨论的甲，乙两射手，甲的方差大，即技术不够“稳定”；而乙射手方差小，较“稳定”，这与习惯看法是一致的.

当然方差也由概率分布完全确定．下面计算一些重要分布的方差，为书写方便，一律假定相应的随机变量为ξ。

[例 11]　二项分布

$$E\xi^2=\sum_{k=0}^{n}k^2\binom{n}{k}p^kq^{n-k}=npq+n^2p^2$$

$$D\xi=E\xi^2-(E\xi)^2=np+n^2p^2-n^2p^2=npq \tag{23}$$

[例 12]　普阿松分布

$$E\xi^2=\sum_{k=0}^{\infty}k^2p_k=\sum_{k=1}^{\infty}k^2\cdot\frac{\lambda^k}{k!}e^{-\lambda}=\sum_{k=1}^{\infty}k\frac{\lambda^k}{(k-1)!}e^{-\lambda}$$

$$=\lambda\sum_{k=0}^{\infty}(k+1)\frac{\lambda^k}{k!}e^{-\lambda}=\lambda^2+\lambda$$

$$D\xi=E\xi^2-(E\xi)^2=\lambda^2+\lambda-\lambda^2=\lambda \tag{24}$$

[例 13]　均匀分布 $U[a,b]$

$$E\xi=\int_a^b x\ \frac{1}{b-a}dx=\frac{b+a}{2} \tag{25}$$

$$E\xi^2=\int_a^b x^2\frac{1}{b-a}dx=\frac{b^2+ab+a^2}{3}$$

$$D\xi=E\xi^2-(E\xi)^2=\frac{b^2+ab+a^2}{3}-\left(\frac{b+a}{2}\right)^2=\frac{(b-a)^2}{12} \tag{26}$$

[例 14]　正态分布 $N(a,\sigma^2)$

$$D\xi=\int_{-\infty}^{\infty}(x-a)^2\frac{1}{\sqrt{2\pi}\sigma}e^{-(x-a)^2/2\sigma^2}dx$$

$$=\frac{\sigma^2}{\sqrt{2\pi}}\int_{-\infty}^{\infty}z^2e^{-z^2/2}dz$$

$$=\frac{\sigma^2}{\sqrt{2\pi}}\left[\left(-ze^{-z^2/2}\right)\Big|_{-\infty}^{\infty}+\int_{-\infty}^{\infty}e^{-z^2/2}dz\right]$$

$$=\frac{\sigma^2}{\sqrt{2\pi}}\sqrt{2\pi}=\sigma^2 \tag{27}$$

这样，我们阐明了正态分布中第二个参数 σ 的概率意义，它就是标准差；而正态分布也由它的数学期望及标准差唯一确定。

下面讨论有关方差的一些性质.

性质1 $D\xi=0$ 的充要条件是 $P\{\xi=C\}=1$，这里C是常数.

性质2 $D(C\xi)=C^2D\xi$, 这里C是常数.

性质3 若 $C\neq E\xi$, 则 $D\xi<E(\xi-C)^2$. (28)

[证明] 因为

$$D\xi=E(\xi-E\xi)^2=E(\xi-C)^2-(C-E\xi)^2 \tag{29}$$

对于随机变量 ξ，若它的数学期望 $E\xi$ 及方差 $D\xi$ 都存在，而且 $D\xi>0$，有时要考虑标准化了的随机变量

$$\xi^*=\frac{\xi-E\xi}{\sqrt{D\xi}}$$

显然 $E\xi^*=0$, $D\xi^*=1$，这正是称 ξ^* 为标准化随机变量的理由.

三、相关系数

对于随机向量 $\xi=(\xi_1, \xi_2, \cdots, \xi_n)$，定义它的方差为$(D\xi_1, D\xi_2, \cdots, D\xi_n)$.

方差反映了随机向量各个分量对于各自的数学期望的离散程度，它对于了解随机向量的分布有一定帮助. 但对于随机向量，我们除了关心它的每个分量的情况外，还希望知道各个分量之间的联系，这光靠数学期望与方差是办不到的. 下面引进的量则能起这个作用，它称为 ξ_i 与ξ_j 的协方差.

$$b_{ij}=\mathrm{cov}(\xi_i, \xi_j)=E(\xi_i-E\xi_i)(\xi_j-E\xi_j),\ i, j=1, 2, \cdots, n \tag{30}$$

显然 $b_{ii}=D\xi_i$，矩阵

$$B=\begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ b_{n1} & b_{n2} & \cdots & b_{nn} \end{pmatrix} \tag{31}$$

称为 ξ 的协方差矩阵，简记作 $D\xi$，显然这是一个对称矩阵.

此外，对任何实数 $t_j(j=1,2,\cdots,n)$ 有

$$\int\cdots\int\left\{\sum_{j=1}^{n}t_j(x_j-E\xi_j)\right\}^2dF(x_1,\cdots,x_n)=\sum_{j,k}b_{jk}t_jt_k\geqslant 0 \quad (32)$$

因此 $\boldsymbol{B}$ 是一个非负定矩阵，所以若以 $|\boldsymbol{B}|$ 记 $\boldsymbol{B}$ 的行列式，则有

$$|\boldsymbol{B}|\geqslant 0 \quad (33)$$

更常用的是如下"标准化"了的协方差：

$$r_{ij}=\frac{\operatorname{cov}(\xi_i,\xi_j)}{\sqrt{D\xi_i}\sqrt{D\xi_j}} \quad (3$$

它称为 ξ_i 与 ξ_j 的**相关系数**，这里当然要求 $D\xi_i$ 与 $D\xi_j$ 不为零。

相关系数也就是标准化的随机变量 $\frac{\xi_i-E\xi_i}{\sqrt{D\xi_i}}$ 与 $\frac{\xi_j-E\xi_j}{\sqrt{D\xi_j}}$ 的协方差.

现在讨论相关系数的性质. 因为相关系数只与两个随机变量有关，下面讨论将假定这两个随机变量是 ξ 与 η，它们的相关系数记为 r.

先证明一条常用的定理.

定理 2（柯西-许瓦兹不等式） 对任意随机变量 ξ 与 η 都有

$$|E\xi\eta|^2\leqslant E\xi^2\cdot E\eta^2 \quad (35)$$

等式成立当且仅当

$$P\{\eta=t_0\xi\}=1 \quad (36)$$

这里 t_0 是某一个常数.

［证明］ 对任意实数 t，定义 $u(t)=E(t\xi-\eta)^2=t^2E\xi^2-2tE\xi\eta+E\eta^2$，显然对一切 t，$u(t)\geqslant 0$，因此二次方程 $u(t)=0$ 或者没有实根或者有一个重根. 所以

$$[E\xi\eta]^2-E\xi^2\cdot E\eta^2\leqslant 0$$

这正是(35). 此外，方程 $u(t)=0$ 有一个重根 t_0 存在的充要条件是

$$[E\xi\eta]^2-E\xi^2E\eta^2=0$$

这时 $E(t_0\xi-\eta)^2=0$, 因此

$$P\{t_0\xi-\eta=0\}=1$$

这就是(36)式. 定理证毕.

由定理 2 立即可以推出, 若两随机变量的方差存在, 则它们的协方差也存在.

把定理 2 应用到随机变量 $\frac{\xi-E\xi}{\sqrt{D\xi}}$ 及 $\frac{\eta-E\eta}{\sqrt{D\eta}}$, 可以得到相关系数的如下重要性质.

性质 1 对相关系数 r, 成立

$$|r|\leqslant 1 \tag{37}$$

并且 $r=1$ 当且仅当

$$P\left\{\frac{\xi-E\xi}{\sqrt{D\xi}}=\frac{\eta-E\eta}{\sqrt{D\eta}}\right\}=1 \tag{38}$$

而 $r=-1$ 当且仅当

$$P\left\{\frac{\xi-E\xi}{\sqrt{D\xi}}=-\frac{\eta-E\eta}{\sqrt{D\eta}}\right\}=1 \tag{39}$$

性质 1 表明, 当 $r=\pm1$ 时, ξ 与 η 存在着线性关系, 这时如果给定一个随机变量之值, 另一个随机变量的值便完全决定.

有线性关系是一个极端, 另一个极端是 $r=0$ 的场合. 为此我们引进:

定义 若随机变量 ξ 与 η 的相关系数 $r=0$, 则我们称 ξ 与 η **不相关**.

性质 2 对随机变量 ξ 与 η, 下面事实是等价的:

(i) $\operatorname{cov}(\xi,\eta)=0$;

(ii) ξ 与 η 不相关;

(iii) $E\xi\eta=E\xi E\eta$;

(iV) $D(\xi+\eta)=D\xi+D\eta$.

[证明] 显然(i)与(ii)是等价的. 由于

$$\mathrm{cov}(\xi,\eta)=E(\xi-E\xi)(\eta-E\eta)=E\xi\eta-E\xi\cdot E\eta \tag{40}$$

因此(i)与(iii)等价. 又由于

$$D(\xi+\eta)=E(\xi-E\xi+\eta-E\eta)^2=D\xi+D\eta+2\,\mathrm{cov}(\xi,\eta) \tag{41}$$

因此(i)与(iV)等价.

独立性和不相关性都是随机变量间联系"薄弱"的一种反映，自然要求知道这两个概念之间的联系. 首先，我们有

性质 3 若 ξ 与 η 独立，则 ξ 与 η 不相关.

[证明] 我们只对连续型随机变量给出证明.

因为 ξ 与 η 独立，故其密度函数 $p(x,y)=p_1(x)p_2(y)$，因此

$$\begin{aligned}\mathrm{cov}(\xi,\eta)&=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-E\xi)(y-E\eta)p(x,y)dxdy\\&=\int_{-\infty}^{\infty}(x-E\xi)p_1(x)dx\cdot\int_{-\infty}^{\infty}(y-E\eta)p_2(y)dy=0\end{aligned}$$

结合性质 2 及性质 3，可以得到，若 ξ 与 η 独立，则 $E\xi\eta=E\xi\cdot E\eta$ 及 $D(\xi+\eta)=D\xi+D\eta$ 成立. 同样的论证可以证明这个结论在 n 个随机变量的场合也成立，即若 $\xi_1,\xi_2,\cdots,\xi_n$ 是相互独立的随机变量，则

$$E\xi_1\xi_2\cdots\xi_n=E\xi_1E\xi_2\cdots E\xi_n \tag{42}$$

$$D(\xi_1+\xi_2+\cdots+\xi_n)=D\xi_1+D\xi_2+\cdots+D\xi_n \tag{43}$$

由独立性可以推出不相关性，但是反过来是不成立的，试看下例.

[例 15] 设 θ 服从 $[0,2\pi]$ 均匀分布，$\xi=\cos\theta$，$\eta=\cos(\theta+a)$，这里 a 是定数. 我们有

$$E\xi=\frac{1}{2\pi}\int_0^{2\pi}\cos t\,dt=0,\ E\eta=\frac{1}{2\pi}\int_0^{2\pi}\cos(t+a)dt=0$$

$$E\xi^2=\frac{1}{2\pi}\int_0^{2\pi}\cos^2t\,dt=\frac{1}{2},\ E\eta^2=\frac{1}{2\pi}\int_0^{2\pi}\cos^2(t+a)dt=\frac{1}{2}$$

$$E\xi\eta=\frac{1}{2\pi}\int_0^{2\pi}\cos t\cos(t+a)\,dt=\frac{1}{2}\cos a$$

因此

$$r=\cos\alpha$$

当 $\alpha=0$ 时, $r=1, \xi=\eta$
当 $\alpha=\pi$ 时, $r=-1, \xi=-\eta$ } 存在线性关系

但是，当 $\alpha=\frac{\pi}{2}$ 或 $\frac{3\pi}{2}$ 时, $r=0$, 这时 ξ 与 η 不相关. 不过, 这时却有 $\xi^2+\eta^2=1$, 因此 ξ 与 η 不独立.

这个例子给我们: (1)提供了 $r=\pm1$ 之例; (2)提供 $r=0$ 之例; (3)说明不能由不相关性推出独立性; (4)说明即使 ξ 与 η 不相关, 它们之间也还是可能存在函数关系. 事实上, 相关系数只是 ξ 与 η 间线性关系程度的一种量度.

不过, 在一种重要的特殊场合——正态分布, 独立性与不相关性却是一致的. 我们先对二维的场合来讨论这个事实.

为此, 我们先求二元正态分布(3.2.22)的相关系数。

$$b_{12}=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-a)(y-b)p(x,y)dxdy$$

$$=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}}\int_{-\infty}^{\infty}e^{-(y-b)^2/2\sigma_2^2}dy\int_{-\infty}^{\infty}(x-a)$$

$$\cdot(y-b)\exp\left\{-\frac{1}{2(1-r^2)}\left(\frac{x-a}{\sigma_1}-r\frac{y-b}{\sigma_2}\right)^2\right\}dx$$

作变数变换 $z=\frac{1}{\sqrt{1-r^2}}\left(\frac{x-a}{\sigma_1}-r\frac{y-b}{\sigma_2}\right), t=\frac{y-b}{\sigma_2}$, 则

$$b_{12}=\frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(\sigma_1\sigma_2\sqrt{1-r^2}\,tz+r\sigma_1\sigma_2t^2)e^{-\frac{t^2}{2}-\frac{z^2}{2}}dzdt$$

$$=\frac{r\sigma_1\sigma_2}{2\pi}\int_{-\infty}^{\infty}t^2e^{-t^2/2}dt\int_{-\infty}^{\infty}e^{-z^2/2}dz+\frac{\sigma_1\sigma_2\sqrt{1-r^2}}{2\pi}$$

$$\cdot\int_{-\infty}^{\infty}te^{-t^2/2}dt\int_{-\infty}^{\infty}ze^{-z^2/2}dz=r\sigma_1\sigma_2$$

此

$$r_{12}=\frac{b_{12}}{\sigma_1\sigma_2}=r \tag{44}$$

这样说明了原来 r 是二元正态分布的相关系数.

至此,我们已经完全搞清了二元正态分布中各个参数的含义. a,b 分别是两个边际分布的数学期望,而 σ_1,σ_2,r 则以下列形式构成它的协方差矩阵

$$\begin{pmatrix} \sigma_1^2 & r\sigma_1\sigma_2 \\ r\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} \tag{45}$$

另外,在上章§2例4中已指出,二元正态分布场合,独立的充要条件是 $r=0$,这表明:

性质 4 对于二元正态分布,不相关性与独立性是等价的。

在§5中,我们将把这个结果推广到多元的场合.

下面,我们给出一个边际分布是正态分布而联合分布不是多元正态分布的例子.

[例16] 令 $\varphi(x)=\frac{1}{\sqrt{2\pi}}e^{-x^2/2}$, $-\infty<x<\infty$,

$$g(x)=\begin{cases}\cos x, & |x|<\pi \\ 0, & |x|\geqslant\pi\end{cases}$$

$$p(x,y)=\varphi(x)\varphi(y)+\frac{1}{2\pi}e^{-x^2}g(x)g(y), \quad -\infty<x,y<\infty$$

关于 $p(x,y)$,不难验证:(1)是二元密度函数;(2)边际分布都是正态分布;(3)相关系数为0;(4)不独立;(5)不是二元正态密度函数. 这些留给读者作为练习.

***四、条件数学期望,最佳线性预测**

在第三章§2我们曾经引进了条件分布函数的概念,现在要相应地引进条件数学期望的概念,并说明它的应用.

为方便起见,我们讨论两个随机变量 ξ 与 η 的场合,假定它们具有密度函数 $p(x,y)$,并以 $p(y|x)$ 记已知 $\xi=x$ 的条件下,η的条

件密度函数，以 $p_1(x)$ 记 ξ 的密度函数。

定义 在 $\xi=x$ 的条件下，η 的条件数学期望定义为

$$E\{\eta|\xi=x\}=\int_{-\infty}^{\infty}yp(y|x)dy \tag{46}$$

［例17］ 若(ξ,η)服从二元正态分布，则由(3.2.29)知

$$p(y|x)=\frac{1}{\sigma_2\sqrt{2\pi}\sqrt{1-r^2}}\exp\left\{-\frac{1}{2\sigma_2^2(1-r^2)}\cdot\left[y-\left(b+r\frac{\sigma_2}{\sigma_1}(x-a)\right)\right]^2\right\}$$

这是正态分布 $N\left(b+r\frac{\sigma_2}{\sigma_1}(x-a),\ \sigma_2^2(1-r^2)\right)$，因此

$$E\{\eta|\xi=x\}=b+r\frac{\sigma_2}{\sigma_1}(x-a) \tag{47}$$

值得提醒的是：这时条件数学期望 $E\{\eta|\xi=x\}$ 是 x 的线性函数.

条件数学期望在预测问题中起重要作用. 问题这样提出：若 ξ，η 是相依的随机变量，我们要找 ξ 与 η 的函数关系，设这个关系是 $y=h(x)$，如果 $E\eta^2$ 及 $E[h(\xi)]^2$都存在，我们的目的是找函数$h(x)$，使 η 与 $h(\xi)$"尽可能靠近"，这里的"靠近"需要一个标准，最常用的是所谓"最小二乘法"，这时要求使

$$E[\eta-h(\xi)]^2 \tag{48}$$

达到最小. 因为

$$\begin{aligned}E[\eta-h(\xi)]^2&=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}[y-h(x)]^2p(x,y)dxdy\\&=\int_{-\infty}^{\infty}p_1(x)\left\{\int_{-\infty}^{\infty}[y-h(x)]^2p(y|x)dy\right\}dx\end{aligned} \tag{49}$$

由(28)知道，当 $h(x)=E\{\eta|\xi=x\}$时，$\int_{-\infty}^{\infty}[y-h(x)]^2p(y|x)dy$ 达到最小，从而使(49)达到最小. 即当我们观察到 $\xi=x$ 时，$E\{\eta|\xi=x\}$是一切对 η 的估值中均方误差最小的一个.

今后我们将称 $y=E\{\eta|\xi=x\}$是 η 关于 ξ 的回归.

通常，(ξ,η)的联合分布函数是不知道的，或者虽然知道但是却不易算出 $E\{\eta|\xi=x\}$．假定已知 ξ 与 η 的数学期望 μ_1,μ_2，标准差 σ_1,σ_2 及相关系数 r，这时可以降低一点要求，改为求**最佳线性预测**．也就是说，把 $h(x)$限定为 x 的线性函数 $L(x)=a+bx$，求 a,b 使

$$e(a,b)=E[\eta-(a+b\xi)]^2 \tag{50}$$

达到最小．

把 $e(a,b)$对 a,b 求偏导数并令它们等于 0，得到

$$\begin{aligned}&2E[\eta-(a+b\xi)]=0\\&2E[(\eta-(a+b\xi))\xi]=0\end{aligned} \tag{51}$$

整理后变成

$$\begin{aligned}&a+b\mu_1=\mu_2\\&a\mu_1+bE\xi^2=E\xi\eta\end{aligned} \tag{52}$$

因此解得

$$a=\mu_2-b\mu_1,\qquad b=\frac{\operatorname{cov}(\xi,\eta)}{\sigma_1^2}=r\cdot\frac{\sigma_2}{\sigma_1} \tag{53}$$

最佳线性预测为

$$L(x)=\mu_2+r\frac{\sigma_2}{\sigma_1}(x-\mu_1) \tag{54}$$

我们称(54)为 η 关于 ξ 的**线性回归**．这个结果与 $E\{\eta|\xi=x\}$一般是不同的，但是在(ξ,η)是二元正态分布的场合，由(47)知两者是重合的，所以在正态分布场合，最佳预测是线性预测，这是一个十分重要的结果．

进一步，我们还可以计算最佳线性预测的均方误差．

$$\begin{aligned}E[\eta-L(\xi)]^2&=E[\eta-\mu_2-b(\xi-\mu_1)]^2\\&=\sigma_2^2+b^2\sigma_1^2-2b\operatorname{cov}(\xi,\eta)\\&=\sigma_2^2-\frac{\operatorname{cov}^2(\xi,\eta)}{\sigma_1^2}=\sigma_2^2(1-r^2)\end{aligned} \tag{55}$$

因此预测误差同 η 的方差有关，也同 ξ 与 η 的相关系数有关，特别当 $|r|=1$ 时(这时 ξ 与 η 有线性关系)，预测误差为 0，也就是说，可以完全准确地进行线性预测．从这个讨论再次看出，相关系数反映了 ξ 与 η 线性联系的程度．

若以 $E\{\eta|\xi\}$ 记随机变量 ξ 的如下函数：当 $\xi=x$ 时，它取值 $E\{\eta|\xi=x\}$．这样定义的 $E\{\eta|\xi\}$ 是随机变量，对它可以求数学期望，并有下列关系式：

$$E\eta=E[E\{\eta|\xi\}] \tag{56}$$

这是条件数学期望的一个极端重要的性质，有广泛应用，在离散型及连续型随机变量的场合，读者不难对它加以证明．

五、矩

数学期望，方差，协方差是随机变量最常用的数字特征，它们都是某种矩．矩是最广泛的一种数字特征，在概率论和数理统计中占有重要地位．最常用的矩有两种：一种是原点矩，对正整数 k，

$$m_k=E\xi^k \tag{57}$$

称为 k 阶原点矩．数学期望是一阶原点矩．

另一种是中心矩，对正整数 k，称

$$c_k=E(\xi-E\xi)^k \tag{58}$$

为 k 阶中心矩．方差是 2 阶中心矩．

由于

$$c_k=E(\xi-E\xi)^k=\sum_{i=0}^{k}\binom{k}{i}(-E\xi)^{k-i}E\xi^i$$

$$=\sum_{i=0}^{k}\binom{k}{i}(-m_1)^{k-i}m_i \tag{59}$$

故中心矩可通过原点矩来表达，反之，

$$m_k=E\xi^k=E[(\xi-m_1)+m_1]^k$$

$$=\sum_{i=0}^{k}\binom{k}{i}E(\xi-m_1)^{k-i}m_1^i=\sum_{i=0}^{k}\binom{k}{i}c_{k-i}m_1^i \tag{60}$$

因此当已知数学期望之后，原点矩也可以通过中心矩给出。

此外对正数 p 还可以定义 p 阶原点绝对矩 $E|\xi|^p$ 及 p 阶中心绝对矩 $E|\xi-E\xi|^p$，它们较少使用。

[例 18] 设 ξ 为正态随机变量，其密度函数为

$$p(x)=\frac{1}{\sqrt{2\pi}\,\sigma}e^{-x^2/2\sigma^2}$$

因此 $E\xi=0$，故

$$m_k=c_k=E\xi^k=\int_{-\infty}^{\infty}x^k\frac{1}{\sqrt{2\pi}\,\sigma}e^{-x^2/2\sigma^2}dx$$

$$=\begin{cases}0, & k\text{ 为奇数}\\ \sigma^k(k-1)(k-3)\cdots3\cdot1, & k\text{ 为偶数}\end{cases} \tag{61}$$

对于多维分布，可以定义各种混合矩，最重要的是协方差。我们不一一赘述了。

*§ 2. 熵 与 信 息①

一、不肯定性与熵

随机试验的主要特征是在试验之前无法肯定地知道哪一个结果将会出现，即随机试验具有一种不肯定性。但是对于不同的随机试验，这种不肯定性的程度却可以有很大的差别。譬如，以射击为例，若有两个射手，他们的射击情况分别以下列两个随机试验来描述：

$$\text{甲}:\begin{pmatrix}A, & \bar{A}\\ 0.5, & 0.5\end{pmatrix}\qquad \text{乙}:\begin{pmatrix}A, & \bar{A}\\ 0.99, & 0.01\end{pmatrix}$$

① 此节内容与本书其余部分基本上独立，初学时可跳过。

这里A表示射中目标，$\bar{A}$表示未射中目标，下一行是相应的概率。显然这两个试验的不肯定性程度很不相同，甲的不肯定性要大得多。假如还有第三个射手，用来描述其射击水平的随机试验为

$$丙：\begin{pmatrix} A, & \bar{A} \\ 0.7, & 0.3 \end{pmatrix}$$

显然应认为此试验的不肯定性程度介于上述二者之间。

因此有必要从数值上估计各种各样随机试验的不肯定性程度，即我们希望找到一个量，用它可以合理地作为不肯定性程度的度量。这样的一个量已经被美国数学家申农(Shannon)找到。

假定我们研究的随机试验α只有有限个不相容的结果A_1，A_2，…，A_n，它们相应的概率为$p(A_1)$，$p(A_2)$，…，$p(A_n)$，满足$\sum_{i=1}^{n} p(A_i)=1$，简写如下：

$$\alpha: \begin{pmatrix} A_1, & A_2, & \cdots, & A_n \\ p(A_1), & p(A_2), & \cdots, & p(A_n) \end{pmatrix}$$

我们希望找到一个量$H(\alpha)$来度量α的不肯定性程度。这个量当然依赖于$p(A_1)$，$p(A_2)$，…，$p(A_n)$，因此亦记为$H(p(A_1), p(A_2), \cdots, p(A_n))$。

为了具体定出$H(\alpha)$的表达式，我们先考察一下，到底$H(\alpha)$应满足什么要求。

首先，我们要求

(i) H是$p(A_i)$的连续函数；

这个要求相当自然，一方面$p(A_i)$的微小变化当然不应引起H的巨大变化，同时也只有连续函数才便于数学上处理。

其次，我们考虑一种特殊的随机试验，这种试验有n个结果，且各个结果出现的概率均为$\frac{1}{n}$，以后简称为有n个等概结果的试

验. 在这种特殊试验中, H 当然应该只是 n 的函数, 并且当 n 增大时, 也即试验有更多可能的结果时, 其相应的不肯定性程度也随之增加.

因此, 我们对 H 提出要求:

(ii) 对有 n 个等概结果的试验, H 是 n 的单调上升函数;

对 $H(\alpha)$的第三个要求比较复杂, 它牵涉到把一个试验分为相继的两个试验. 我们通过简单的例子来阐述其含义.

考虑有三个结果的试验

$$\alpha: \begin{pmatrix} A_1, A_2, A_3 \\ p_1, p_2, p_3 \end{pmatrix}$$

这个试验的不肯定性程度 $H(\alpha)=H(p_1, p_2, p_3)$. 为了确定到底是哪一个结果出现, 我们也可以进行这样相继的两个试验: 在第一个试验中, 先确定到底是 A_1 出现, 还是 A_2 或 A_3 出现, 即进行下列试验:

$$\alpha_1: \begin{pmatrix} A_1, & B \\ p_1, & p_2+p_3 \end{pmatrix}$$

显然 $H(\alpha_1)=H(p_1, p_2+p_3)$. 如果 A_1 出现(其概率为 p_1), 则试验结果已完全确定, 无须再作进一步的试验; 但是如果是 B 出现(其概率为 p_2+p_3), 则尚须进行如下试验, 才能最后确定试验结果,

$$\alpha_2: \begin{pmatrix} A_2, & A_3 \\ \dfrac{p_2}{p_2+p_3}, & \dfrac{p_3}{p_2+p_3} \end{pmatrix}$$

这个试验的不肯定性程度为 $H(\alpha_2)=H\left(\dfrac{p_2}{p_2+p_3}, \dfrac{p_3}{p_2+p_3}\right)$.

可以直接进行试验 α, 以确定 A_1, A_2, A_3 中哪一个结果出现; 但是若先进行试验 α_1, 然后有必要时(概率 p_2+p_3)再进行 α_2, 也可达到同样的目的. 因此, 我们自然地认为这两组试验所含的不肯定性程度是一样的, 即

$$H(p_1, p_2, p_3)=H(p_1, p_2+p_3)+(p_2+p_3)H\left(\frac{p_2}{p_2+p_3}, \frac{p_3}{p_2+p_3}\right)$$

这些考虑启发我们对H提出下列要求:

(iii) 一个试验分成相继的两个试验时,未分之前的H是既分之后的H的加权和.

条件(i)(ii)(iii)已完全确定H的形式.下面简记$p(A_i)$为p_i。

定理(申农) 唯一满足 (i)、(ii)、(iii) 三个条件的H具有下列形式

$$H=-C\sum_{i=1}^{n}p_i\log p_i \tag{1}$$

其中C是正常数.

为证明这个定理,要用到下述分析引理.

引理1 若$f(n)$是n的单调上升函数,且对一切正整数m,n成立

$$f(mn)=f(m)+f(n) \tag{2}$$

则

$$f(n)=C\log n$$

其中C是一个正常数.

[证明] 由(2)可得$f(1)=0$, 所以对其他正整数m, 有$f(m)>0$. 另一方面,

$$f(n^2)=f(n)+f(n)=2f(n)$$
$$f(n^3)=f(n^2)+f(n)=3f(n)$$

一般地

$$f(n^k)=kf(n) \tag{3}$$

若n, m是两个任意的正整数, $m\neq1$,选任意大的正整数k, 再取正整数l,使

$$m^l\leqslant n^k<m^{l+1} \tag{4}$$

由函数的单调性

$$f(m^l)\leqslant f(n^k)<f(m^{l+1})$$

由(3)得

$$lf(m)\leqslant kf(n)<(l+1)f(m)$$

因此

$$\frac{l}{k}\leqslant\frac{f(n)}{f(m)}<\frac{l+1}{k}$$

对(4)取对数得

$$l\log m\leqslant k\log n<(l+1)\log m$$

因此也有

$$\frac{l}{k}\leqslant\frac{\log n}{\log m}<\frac{l+1}{k}$$

这样一来

$$\left|\frac{f(n)}{f(m)}-\frac{\log n}{\log m}\right|<\frac{1}{k}$$

上式对任意大的 k 都成立,因此

$$\frac{f(n)}{f(m)}=\frac{\log n}{\log m}$$

由于 m,n 的任意性即知

$$f(n)=C\log n$$

其中 C 是常数,再由 $f(n)$ 是 n 的上升函数,可知 C 是正的。

现在可以来证明定理了.

[证明] 首先,记 $H\left(\frac{1}{n},\frac{1}{n},\cdots,\frac{1}{n}\right)=f(n)$, 按条件(ii)知 $f(n)$ 是 n 的单调上升函数. 对有 mn 个等概结果的试验,可以把它分解为 m 个有 n 个等概结果的试验,因此由条件(iii)知

$$f(mn)=f(m)+m\cdot\frac{1}{m}f(n)=f(m)+f(n)$$

利用引理1立刻得到

$$H\left(\frac{1}{n}, \frac{1}{n}, \cdots, \frac{1}{n}\right)=C\log n$$

其次，当 $p_1, p_2, \cdots, p_n$ 是有理数时，不妨记 $p_i=\frac{n_i}{\sum_{i=1}^{n} n_i}$，考虑一个有 $\sum_{i=1}^{n} n_i$ 个等概结果的试验，而这个试验又可以看作两个相继的试验，其中第一个试验以概率 p_i 出现结果 A_i；而第二个试验，则是在出现结果 A_i 的基础上，考察它是出现 n_i 个等概结果中的哪一个，因此按条件(iii)应有

$$C\log\sum_{i=1}^{n} n_i=H(p_1, p_2, \cdots, p_n)+C\sum_{i=1}^{n} p_i\log n_i$$

于是

$$\begin{aligned}H(p_1, p_2, \cdots, p_n)&=C\left[\log\sum_{i=1}^{n} n_i-\sum_{i=1}^{n} p_i\log n_i\right]\\&=C\left[\sum_{i=1}^{n} p_i\left(\log\sum_{i=1}^{n} n_i-\log n_i\right)\right]\\&=-C\sum_{i=1}^{n} p_i\log p_i\end{aligned}$$

最后，对任意的 $p_1, p_2, \cdots, p_n$，可用有理数来逼近它，但按条件(i)知H是各自变量的连续函数，因此上述表达式仍然成立，从而完成了定理证明。

定理中的系数 C，可以根据方便选择，它取决于度量单位。常用的度量单位有二进制单位及十进制单位，前者对数的底取为 2，后者用常用对数。

在(1)中，若 $p_i=0$，则相应的项 $p_i\log p_i$ 定义为零，因此在试验中增减零概率结果不影响不肯定性，这是很自然的。

以后将称

$$H(\alpha)=-\sum_{i=1}^{n}p(A_i)\log p(A_i) \tag{5}$$

为试验 α 的熵.

从下面例子可以看出,熵确实可以度量试验的不肯定性程度.

[例 1] 计算本节开始时射击例子中射手甲,乙,丙射击试验相应的熵.

$$H_{甲}=-\frac{1}{2}\lg\frac{1}{2}-\frac{1}{2}\lg\frac{1}{2}=\lg 2=0.3010$$

$$H_{乙}=-0.99\lg 0.99-0.01\lg 0.01=0.0243$$

$$H_{丙}=-0.7\lg 0.7-0.3\lg 0.3=0.2653$$

这里均用十进制单位. 甲的熵最大,乙最小,丙介于其中,与直观完全符合.

[例 2] 英文字母出现的熵.

熵的概念的引入与通信理论的发展密切相关,后者需要解决最有效而无错误地传递消息这一任务. 用文字给出的消息通常经过编码变成某种信号,经信道传输,到达接收地点,再经译码还原. 当然编码的优劣将直接影响通信的效率. 最简单的办法是把每一个字母都转换成相同长度的码子,但正如本书一开头就看到的那样,在英文中不同字母出现的频率很不相同. 因此更有效的办法当然是把最常出现的字母编成较短的码子,而把不常出现的字母编成较长的码子. 这个问题与英文字母出现的熵有直接关系.

假如把 26 个字母连同分隔用的空格共 27 个符号,看作是等可能出现的,则相应的试验(接收到一个符号,要判断它是哪个字母或空格)的熵为

$$H_0=\lg 27=1.4314$$

但是接收到一个符号,它是空格的可能性比是 z 的可能性要大200

倍，也即不肯定性程度似乎不应该有这么大．事实上，如果我们考虑到不同字母及空格的出现概率（数值见第一章 § 1），则

$$H_1=-\sum_{i=1}^{27}p_i\lg p_i=1.213$$

也就是说不肯定性要小不少．当然如果再注意到英文中前后字母间的联系，例如 q 总是接着 u，则这种不肯定性还要小得多．

二、熵的基本性质

下面讨论用(5)定义的熵的若干基本性质，通过这些研究将进一步看到熵作为不肯定性程度度量的合理性．

首先考察一下函数 $\varphi(x)=-x\log x$ 的性质，显然对于 $x>0$，均有 $\varphi''(x)<0$，因此 $\varphi(x)$ 是 $(0,\infty)$ 上的上凸函数，即对于任意的 $p>0, q>0$，且 $p+q=1$，不等式

$$p\varphi(x_1)+q\varphi(x_2)<\varphi(px_1+qx_2) \tag{6}$$

对一切 $0<x_1<x_2<\infty$ 成立．

一般地，不难用归纳法证明如下分析引理．

引理 2（Jensen 不等式） 设 $\varphi(x)$ 是 $[a, b]$ 上的上凸函数，而 $x_1, x_2, \cdots, x_n$ 是 $[a, b]$ 中的任意点，$\lambda_1, \lambda_2, \cdots, \lambda_n$ 是和为 1 的正数，则

$$\sum_{i=1}^{n}\lambda_i\varphi(x_i)\leqslant\varphi\left(\sum_{i=1}^{n}\lambda_i x_i\right) \tag{7}$$

等号成立当且仅当诸 x_i 相等．

下面证明熵的若干性质．

1．当且仅当 $p(A_i)$，$i=1, 2, \cdots, n$ 之中的一个等于 1 时，熵 $H=0$，其他情况下，熵恒为正．

2．在有 n 个可能结果的试验中，等概试验具有最大熵，其值为 $\log n$．

[证明] 在引理 2 中取 $\varphi(x)=-x\log x$, $x_i=p(A_i)$, $\lambda_i=\frac{1}{n}$, 代入(7)式, 得到

$$\frac{-1}{n}\sum_{i=1}^{n}p(A_i)\log p(A_i)\leqslant\frac{-1}{n}\log\frac{1}{n}$$

即

$$H(p(A_1),\cdots,p(A_n))\leqslant\log n=H\left(\frac{1}{n},\cdots,\frac{1}{n}\right)$$

下面考虑两个试验 α 及 β, 设它们的结果及概率如下

$$\alpha:\begin{pmatrix}A_1, & \cdots, & A_m\\ p(A_1), & \cdots, & p(A_m)\end{pmatrix}\qquad \beta:\begin{pmatrix}B_1, & \cdots, & B_n\\ p(B_1), & \cdots, & p(B_n)\end{pmatrix}$$

又以 $\alpha\beta$ 记这两个试验联合起来所构成的新试验, 于是试验 $\alpha\beta$ 的可能结果为 A_kB_l, $k=1, 2, \cdots, m$, $l=1, 2, \cdots, n$, 相应的概率为 $p(A_kB_l)$. 按定义

$$H(\alpha\beta)=-\sum_{k,l}p(A_kB_l)\log p(A_kB_l) \tag{8}$$

3. 若试验 α 与试验 β 独立, 则

$$H(\alpha\beta)=H(\alpha)+H(\beta) \tag{9}$$

[证明] 在这种场合 $p(A_kB_l)=p(A_k)p(B_l)$, 因此

$$\begin{aligned}H(\alpha\beta)&=-\sum_{k,l}p(A_k)p(B_l)\log p(A_k)p(B_l)\\&=-\sum_{k,l}p(A_k)p(B_l)[\log p(A_k)+\log p(B_l)]\\&=H(\alpha)+H(\beta)\end{aligned}$$

三、条件熵与信息量

为了进一步研究熵的性质, 需要引进条件熵的概念. 设 α, β 是前述两个试验, 以 $p(B_l|A_k)$ 记试验 α 出现结果 A_k 的条件下, 试验 β 出现结果 B_l 的概率, 则

$$H_{A_k}(\beta)=-\sum_{l=1}^{n}p(B_l|A_k)\log p(B_l|A_k) \tag{10}$$

是在试验 α 出现 A_k 的条件下，试验 β 的熵。

我们称平均值

$$H_\alpha(\beta)=\sum_{k=1}^{m}p(A_k)H_{A_k}(\beta) \tag{11}$$

为在试验 α 实现的条件下试验 β 的**条件熵**。

下面指出 $H_\alpha(\beta)$ 的某些重要性质。

1. $H(\alpha\beta)=H(\alpha)+H_\alpha(\beta)$. (12)

[证明] $H(\alpha\beta)=-\sum_{k,l}p(A_kB_l)\log p(A_kB_l)$

$$=-\sum_{k,l}p(A_k)p(B_l|A_k)[\log p(A_k)+\log p(B_l|A_k)]$$

$$=-\sum_k p(A_k)\log p(A_k)\sum_l p(B_l|A_k)$$

$$-\sum_k p(A_k)\sum_l p(B_l|A_k)\log p(B_l|A_k)$$

$$=H(\alpha)+H_\alpha(\beta)$$

特别当 α 与 β 独立时，$H_\alpha(\beta)=H(\beta)$，此时(12)化为(9)式。

同理，

$$H(\alpha\beta)=H(\beta)+H_\beta(\alpha)$$

这个性质称为**熵的加法法则**．推导熵的表达式时的条件(iii)，事实上是加法法则的另一种表述．

2. $H_\alpha(\beta)$ 是非负的．又若所有的 $p(A_i)>0$，则当且仅当 $H_{A_i}(\beta)=0(i=1,\cdots,m)$时，$H_\alpha(\beta)=0$ 才成立，此时还有 $H(\alpha\beta)=H(\alpha)$．

这些性质用条件熵的定义立即得到．后一结论说明，只有当试验 α 的任何结果都使试验 β 的不肯定性完全消除时，才有 $H_\alpha(\beta)$

$=0$, 此时 α 的结果完全决定了 β 的结果。

3. $H_\alpha(\beta) \leqslant H(\beta)$. (13)

［证明］ 在引理2中，取 $\varphi(x) = -x\log x$, $\lambda_i = p(A_i)$, $x_i = p(B_k|A_i)$, 则

$$-\sum_{i=1}^{m} p(A_i)p(B_k|A_i)\log p(B_k|A_i)$$

$$\leqslant -\left[\sum_{i=1}^{m} p(A_i)p(B_k|A_i)\right]\log\left[\sum_{i=1}^{m} p(A_i)p(B_k|A_i)\right]$$

$$= -p(B_k)\log p(B_k)$$

两边对 k 求和即得

$$-\sum_{i=1}^{m} p(A_i)\sum_{k=1}^{n} p(B_k|A_i)\log p(B_k|A_i) \leqslant -\sum_{k=1}^{n} p(B_k)\log p(B_k)$$

此即

$$H_\alpha(\beta) \leqslant H(\beta)$$

此外，显然有

$$H(\alpha\beta) = H(\alpha) + H_\alpha(\beta) \leqslant H(\alpha) + H(\beta)$$

关系式(13)的涵义不难理解，因为进行试验 α 之后，一般对试验 β 的结果会增加了解，从而消除了部分不肯定性，只有当 α 与 β 独立时，$H_\alpha(\beta) = H(\beta)$，此时 α 的结果无助于减少 β 的不肯定性。因此量 $H(\beta) - H_\alpha(\beta)$ 是作了辅助试验 α 之后试验 β 不肯定性的减少量，即是由于试验 α 的进行而得到的有关试验 β 的信息。

记

$$I(\alpha, \beta) = H(\beta) - H_\alpha(\beta) \tag{14}$$

并称之为含在试验 α 中的有关试验 β 的**信息量**。

因为

$$H(\alpha\beta) = H(\alpha) + H_\alpha(\beta) = H(\beta) + H_\beta(\alpha)$$

所以

$$H(\alpha)-H_{\beta}(\alpha)=H(\beta)-H_{\alpha}(\beta)$$

即

$$I(\beta,\alpha)=I(\alpha,\beta)$$

因此β中含有α的信息量与α中含有β的信息量相等.

显然

$$0\leqslant I(\alpha,\beta)\leqslant H(\alpha),\quad 0\leqslant I(\alpha,\beta)\leqslant H(\beta) \tag{15}$$

当且仅当α与β独立时,才有$I(\alpha,\beta)=0$,此时一个试验不含有另一个试验的任何信息量. 另一个极端情形是当α的结果完全决定β的结果,此时$H_{\alpha}(\beta)=0$,从而$I(\alpha,\beta)=H(\beta)$,特别地,$I(\beta,\beta)=H(\beta)$,这就是说,包含在试验β中有关试验β的信息量等于β的熵,因而熵也是信息量.

四、连续型分布的熵

对于具有密度函数的连续型分布,可以类似地定义它的熵. 设随机变量α及β的密度函数分别为$p(x)$及$q(y)$,它们的联合密度函数为$f(x,y)$. 一种比较显然的定义熵的办法是仿照离散型场合,定义

$$H(\alpha)=-\int_{-\infty}^{\infty}p(x)\log p(x)dx \tag{16}$$

$$H(\alpha\beta)=-\iint f(x,y)\log f(x,y)dxdy \tag{17}$$

和

$$H_{\alpha}(\beta)=-\iint f(x,y)\log\frac{f(x,y)}{p(x)}dxdy \tag{18}$$

$$H_{\beta}(\alpha)=-\iint f(x,y)\log\frac{f(x,y)}{q(y)}dxdy \tag{19}$$

这样定义的熵及条件熵具有许多离散型分布的熵的性质. 简单罗列如下:

1. 若α限制在V中变化，则V中的均匀分布有最大熵，其值等于$\log|V|$，此处$|V|$是V的测度.

2. $H(\alpha\beta)=H(\alpha)+H_\alpha(\beta)=H(\beta)+H_\beta(\alpha)$.

而且

$$H_\alpha(\beta)\leqslant H(\beta)$$

因而

$$H(\alpha\beta)\leqslant H(\alpha)+H(\beta)$$

这两个等号成立的充要条件是α与β独立.

第一个性质说明，从熵的观点看，均匀分布具有特殊地位. 令人惊奇的是，常见的另外两个分布——正态分布与指数分布，从熵的观点看，也具有特殊地位.

3. 设$p(x)$是一维密度函数，其标准差为σ，则当$p(x)$为正态分布时其熵最大，其值等于$\log\sqrt{2\pi e}\,\sigma$(对数以$e$为底).

[证明] 不失一般性，设其数学期望为零，这时要求$p(x)$满足约束条件

$$\int p(x)dx=1$$

及

$$\sigma^2=\int x^2p(x)dx$$

又使

$$H(x)=-\int p(x)\log p(x)dx$$

达到最大. 根据变分法，这相当于要求

$$\int[-p(x)\log p(x)+\lambda p(x)+\mu x^2p(x)]dx$$

达到极大，即

$$-1-\log p(x)+\lambda+\mu x^2=0$$

选取常数使其满足约束条件，即得

$$p(x)=\frac{1}{\sqrt{2\pi}\,\sigma}e^{-x^2/2\sigma^2}$$

此时

$$H(x)=\int\frac{1}{\sqrt{2\pi}\,\sigma}e^{-x^2/2\sigma^2}\left[\log\sqrt{2\pi}\,\sigma+\frac{x^2}{2\sigma^2}\right]dx$$

$$=\log\sqrt{2\pi}\sigma+\frac{1}{2}=\log\sqrt{2\pi e}\,\sigma$$

在 n 维场合，假定分布的协方差矩阵固定为 $B=(b_{ij})$，则

$$p(x)=\frac{1}{(2\pi)^{n/2}|B|^{\frac{1}{2}}}\exp\left\{-\frac{1}{2}xB^{-1}x'\right\},x\in R^n$$

达到最大熵 $\log(2\pi e)^{n/2}|B|^{\frac{1}{2}}$，这里 $|B|$ 为 B 的行列式值.

4. 若密度函数 $p(x)$ 当 $x\leqslant 0$ 时等于 0，并且其均值为 a，则指数分布

$$p(x)=\frac{1}{a}e^{-x/a}$$

达到最大熵，其值为 $\log ea$.

证明方法与推导正态分布时完全一样，不再写出.

连续熵的性质也有与离散熵不同的，特别是它的数值会因坐标系的改变而改变. 因此还存在着别种关于连续熵的定义，不过这里不准备再深入讨论了.

熵是一门新兴学科——信息论中的基本概念. 它的引入使得人们能对随机现象的不肯定性进行度量，是具有重大意义的.

*§3. 母 函 数

一、整值随机变量与母函数的定义

在离散型随机变量中，那些只取非负整数值 0, 1, 2, … 的占有重要的地位. 事实上，我们所遇到的离散型分布如二项分布，超几何分布，普阿松分布，几何分布等都是取非负整数值的.

我们称取非负整数值的随机变量为**整值随机变量**. 对于整值随机变量，有一种处理方法很便于应用，这就是母函数法.

若随机变量 ξ 取非负整数值，且相应的分布列为

$$\begin{pmatrix} 0, & 1, & 2, \cdots \\ p_0, & p_1, & p_2, \cdots \end{pmatrix} \tag{1}$$

则

$$P(s)=\sum_{k=0}^{\infty} p_k s^k \tag{2}$$

称为 ξ 的**母函数**，即

$$P(s)=Es^{\xi} \tag{3}$$

因为母函数由分布列完全决定，因此亦称它为该概率分布的母函数. 由于

$$\sum_{k=0}^{\infty} p_k=1 \tag{4}$$

由幂级数的收敛性知道 $P(s)$ 至少在 $|s| \leqslant 1$ 一致收敛且绝对收敛. 因此母函数对任何整值随机变量都存在.

对于任一数列 $\{a_n\}$ 也可定义 $\sum\limits_{n=0}^{\infty} a_n s^n$ 为其母函数，但我们以后只讨论概率分布对应的母函数.

母函数在十九世纪初被拉普拉斯引进，它是在概率论中第一个被系统地应用的变换法，对后来在概率论中引进其他更有用的变换——如下节要介绍的特征函数——有启发作用. 本书把变换法的重点放在特征函数法上，至于为什么要单独介绍母函数，首先是由于它比较简单，在整值随机变量场合很有用，可以作为特征函数的前导；其次，在随机过程中要用到有关结果；最后还由于从母函数法发展起来的 Z 变换法已成为解决许多问题的重要工具. 但是，跳过本节母函数的内容，并不太影响本书以后章节的学习.

下面求几种分布的母函数。

[例 1] 二项分布

$$P(s)=\sum_{k=0}^{n}\binom{n}{k}p^k q^{n-k}s^k=(q+ps)^n \tag{5}$$

[例 2] 超几何分布

$$P(s)=\sum_{k=0}^{n}\frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}s^k \tag{6}$$

这是超几何级数，是一种特殊函数，处理起来不大方便，在概率论中也很少用，由于超几何分布的名称来自于它，因此我们顺便提及。

[例 3] 普阿松分布

$$P(s)=\sum_{k=0}^{\infty}\frac{\lambda^k}{k!}e^{-\lambda}s^k=e^{-\lambda}\cdot e^{\lambda s}=e^{\lambda(s-1)} \tag{7}$$

[例 4] 几何分布

$$P(s)=\sum_{k=1}^{\infty}q^{k-1}ps^k=ps\sum_{k=1}^{\infty}(qs)^{k-1}=\frac{ps}{1-qs} \tag{8}$$

二、母函数的性质

1．唯一性．由分布列(1)用(2)定义母函数，这显然是唯一确定的；下面证明，由母函数也能唯一确定分布列．

设概率分布$\{p_k\}$及$\{q_k\}$分别具有母函数$P(s)$及$Q(s)$，且$P(s)=Q(s)$，因为$P(s)$及$Q(s)$都是幂级数，且当$|s|\leqslant 1$时收敛，对$P(s)$及$Q(s)$求导k次，并令$s=0$，则得

$$k!p_k=P^{(k)}(0)=Q^{(k)}(0)=k!q_k$$

因此$p_k=q_k$，$k=0,1,2,\cdots$，即两个概率分布一样．

这样一来，概率分布与母函数是一一对应的，因而对于概率分

布的许多研究可以化为对所对应的母函数的研究，因为母函数是幂级数，具有许多良好的性质，便于处理，所以母函数是研究整值随机变量的有效工具.

2. 母函数与数字特征. 母函数的应用之一是利用它能求得概率分布的数字特征，若

$$P(s)=\sum_{k=0}^{\infty}p_k s^k$$

则

$$P'(s)=\sum_{k=1}^{\infty}kp_k s^{k-1},\quad P''(s)=\sum_{k=2}^{\infty}k(k-1)p_k s^{k-2} \tag{9}$$

这两个级数至少在$|s|<1$是收敛的.

当数学期望$\sum_{k=1}^{\infty}kp_k$存在时，

$$P'(1)=\sum_{k=1}^{\infty}kp_k=E\xi \tag{10}$$

当数学期望$\sum_{k=1}^{\infty}kp_k=\infty$时，$\lim_{s\to 1}P'(s)=\infty$.

同样，当方差$D\xi$存在时，

$$E[\xi(\xi-1)]=\sum_{k=2}^{\infty}k(k-1)p_k=P''(1) \tag{11}$$

故

$$D\xi=E\xi^2-(E\xi)^2=P''(1)+P'(1)-[P'(1)]^2 \tag{12}$$

公式(10)及(12)是计算数学期望及方差的简便公式.

[例 5] 二项分布：母函数为$P(s)=(q+ps)^n$.

$$E\xi=P'(1)=n(q+ps)^{n-1}p|_{s=1}=np$$

$$P''(1)=n(n-1)(q+ps)^{n-2}p^2|_{s=1}=n(n-1)p^2$$

$$D\xi = n^2p^2 - np^2 + np - n^2p^2 = npq$$

这结果在§1曾直接计算过.

［例6］ 普阿松分布：母函数为$P(s)=e^{\lambda(s-1)}$.

$$E\xi = P'(1) = e^{\lambda(s-1)}\cdot\lambda|_{s=1} = \lambda$$

$$P''(1) = e^{\lambda(s-1)}\cdot\lambda^2|_{s=1} = \lambda^2$$

$$D\xi = \lambda^2 + \lambda - \lambda^2 = \lambda$$

这结果在§1也直接计算过，这里的计算较方便.

三、独立随机变量和的母函数

若随机变量ξ与η相互独立，它们都是整值随机变量，概率分布分别为$\{a_k\}$及$\{b_k\}$，而相应的母函数为$A(s)$及$B(s)$，下面计算随机变量$\zeta=\xi+\eta$的概率分布. 显然ζ也是整值随机变量，若记$c_r=P\{\zeta=r\}$，则

$$c_r = a_0b_r + a_1b_{r-1} + \cdots + a_rb_0 \tag{13}$$

这就是离散褶积公式(3.3.5).

记

$$C(s) = \sum_{r=0}^{\infty} c_r s^r \tag{14}$$

利用母函数在$|s|\leqslant 1$的一致收敛性及绝对收敛性，

$$A(s)B(s) = \sum_{k=0}^{\infty} a_k s^k \cdot \sum_{l=0}^{\infty} b_l s^l = \sum_{k,l} a_k b_l s^{k+l}$$

$$= \sum_{r=0}^{\infty}\left(\sum_{k=0}^{r} a_k b_{r-k}\right)s^r = \sum_{r=0}^{\infty} c_r s^r$$

因此

$$C(s) = A(s)B(s) \tag{15}$$

即两个独立随机变量之和的母函数是这两个随机变量的母函数的乘积. 这是一个相当重要的性质，由于母函数具有这个性质，因此

在研究独立随机变量和的问题时，母函数很适用.

容易把上面结果推广到 n 个独立整值随机变量之和的场合，若随机变量 $\xi_1, \xi_2, \cdots, \xi_n$ 相互独立，它们的母函数分别为 $P_1(s)$, $P_2(s), \cdots, P_n(s)$, 则 $\eta=\xi_1+\xi_2+\cdots+\xi_n$ 的母函数为

$$P(s)=P_1(s)P_2(s)\cdots P_n(s) \tag{16}$$

特别当 ξ_i 有相同概率分布的场合，$P_i(s)=P_1(s)$, 这时

$$P(s)=[P_1(s)]^n \tag{17}$$

[例 7] 二项分布的母函数：在成功概率为 p 的 n 次贝努里试验中，若令

$$\xi_i=\begin{cases}1, & \text{在第 } i \text{ 次试验中 } A \text{ 出现}\\ 0, & \text{在第 } i \text{ 次试验中 } A \text{ 不出现}\end{cases} \tag{18}$$

则 $\xi_1, \xi_2, \cdots, \xi_n$ 相互独立，而且 $\eta=\xi_1+\xi_2+\cdots+\xi_n$ 服从二项分布. ξ_i 的母函数为$(q+ps)$, 由(17), η 的母函数为

$$P(s)=(q+ps)^n$$

这与例 1 直接计算的结果相同.

[例 8] 从装有号码为 1, 2, 3, 4, 5, 6 的小球的袋中有放回地取 5 个小球，求所得号码数总和为 15 的概率.

[解] 若以 ξ_i 记第 i 次摸得的数字，则总和 $\eta=\xi_1+\xi_2+\cdots+\xi_5$, ξ_i 的母函数为

$$P_i(s)=\frac{1}{6}(s+s^2+s^3+s^4+s^5+s^6)$$

显然 $\xi_1, \xi_2, \cdots, \xi_5$ 是相互独立的，因此 η 的母函数为

$$P(s)=\frac{1}{6^5}(s+s^2+\cdots+s^6)^5$$

所求的概率 $P\{\eta=15\}$ 是 $P(s)$ 展开式中 s^{15} 项的系数. 由于

$$P(s)=\frac{s^5}{6^5}(1+s+\cdots+s^5)^5=\frac{s^5}{6^5}\left(\frac{1-s^6}{1-s}\right)^5$$

$$=\frac{s^5}{6^5}(1-s^6)^5(1-s)^{-5}$$

$$=\frac{s^5}{6^5}(1-5s^6+10s^{12}+\cdots-s^{30})\left[\sum_{k=0}^{\infty}\binom{-5}{k}(-s)^k\right]$$

$$=\frac{s^5}{6^5}(1-5s^6+\cdots)\left[\sum_{k=0}^{\infty}(-1)^k\binom{5+k-1}{k}(-1)^k s^k\right]$$

$$=\frac{s^5}{6^5}(1-5s^6+\cdots)\left(\sum_{k=0}^{\infty}\binom{k+4}{k}s^k\right)$$

故

$$P\{\eta=15\}=\frac{1}{6^5}\left[1\times\binom{14}{10}-5\times\binom{8}{4}\right]=\frac{651}{6^5}$$

四、随机个随机变量之和的母函数

若 $\xi_1,\xi_2,\cdots,\xi_n,\cdots$是一串相互独立具有相同概率分布的整值随机变量，$P\{\xi_i=j\}=f_j$，其母函数为

$$F(s)=\sum_{j=0}^{\infty}f_j s^j$$

随机变量ν是取正整数值的，且 $P\{\nu=n\}=g_n$，其母函数为

$$G(s)=\sum_{n=1}^{\infty}g_n s^n$$

若 $\{\xi_n\}$ 与 ν 独立，考虑和 $\eta=\xi_1+\xi_2+\cdots+\xi_\nu$，记

$$P\{\eta=i\}=h_i$$

我们来求 η 的母函数

$$H(s)=\sum_{i=0}^{\infty}h_i s^i$$

利用全概率公式及$\{\xi_n\}$与 ν 的独立性

$$h_i=P\{\eta=i\}=\sum_{n=1}^{\infty}P\{\nu=n\}P\{\eta=i\,|\,\nu=n\}$$

$$=\sum_{n=1}^{\infty}P\{\nu=n\}P\{\xi_1+\xi_2+\cdots+\xi_n=i\,|\,\nu=n\}$$

$$=\sum_{n=1}^{\infty}P\{\nu=n\}P\{\xi_1+\xi_2+\cdots+\xi_n=i\}$$

由于 $\xi_1+\xi_2+\cdots+\xi_n$ 为 n 个相互独立相同分布的随机变量之和，故其母函数

$$\sum_{i=0}^{\infty}P\{\xi_1+\cdots+\xi_n=i\}s^i=[F(s)]^n$$

因此

$$H(s)=\sum_{i=0}^{\infty}h_i s^i=\sum_{n=1}^{\infty}P\{\nu=n\}\sum_{i=0}^{\infty}P\{\xi_1+\cdots+\xi_n=i\}s^i$$

$$=\sum_{n=1}^{\infty}g_n[F(s)]^n=G[F(s)] \tag{19}$$

从(19)可以看到，随机个相互独立相同分布的随机变量之和的母函数是原来两个母函数的复合.

由于

$$H'(s)=G'[F(s)]\cdot F'(s)$$

因此当 $E\xi_i$ 及 $E\nu$ 存在时，在上式中令 $s=1$，得到

$$E\eta=E\nu\cdot E\xi_i \tag{20}$$

这个公式有不少用处.

[复合普阿松分布] 在上述讨论中，若 ν 服从参数为 λ 的普阿松分布，则

$$G(s)=e^{\lambda(s-1)}$$

因此 $\eta=\xi_1+\xi_2+\cdots+\xi_\nu$ 的母函数为

$$H(s)=e^{\lambda(F(s)-1)} \tag{21}$$

以(21)为母函数的概率分布称为**复合普阿松分布**. 这种分布也很有用.

特别当 $F(s)=q+ps$ 时，

$$H(s)=e^{\lambda p(s-1)} \tag{22}$$

[例 9] 观察资料表明，天空中星体服从普阿松分布，其参数为 λV，这里 V 是被观察区域的体积. 若每个星球上有生命存在的概率为 p，则在体积为 V 的宇宙空间中有生命存在的星球数服从参数为 λpV 的普阿松分布.

§4. 特征函数

一、定义

数字特征只反映了概率分布的某些侧面，一般并不能通过它们来完全确定分布函数，本节将要引进的特征函数，既能完全决定分布函数而又具有良好的分析性质.

为了定义特征函数，我们需要稍许拓广一下随机变量的概念，引进复随机变量.

定义 如果 ξ 与 η 都是概率空间 $(\Omega, \mathscr{F}, P)$ 上的实值随机变量，则称 $\zeta=\xi+i\eta$ 为**复随机变量**.

从定义知道，对复随机变量的研究本质上是对二维随机向量的研究. 这里举一个例子：如果二维向量 (ξ_1, η_1) 与 (ξ_2, η_2) 是独立的，则我们称复随机变量 $\zeta_1=\xi_1+i\eta_1$ 与 $\zeta_2=\xi_2+i\eta_2$ 是独立的.

定义一个复随机变量 $\zeta=\xi+i\eta$ 的数学期望为

$$E\zeta=E\xi+iE\eta$$

对复随机变量也可以平行于实随机变量建立起一系列结果. 例如，若 $\zeta_1, \zeta_2, \cdots, \zeta_n$ 是相互独立的，则

$$E\zeta_1\zeta_2\cdots\zeta_n=E\zeta_1 E\zeta_2\cdots E\zeta_n$$

又如，若 $g(x)$ 是一个一元波雷尔可测函数，而 $\eta=g(\xi)$，则

$$Ee^{it\eta}=Ee^{itg(\xi)}=\int_{-\infty}^{\infty} e^{itg(x)}dF_\xi(x) \tag{1}$$

这里 $e^{it\eta}=\cos t\eta+i\sin t\eta$.

以后将随时引用这类结果而不再加以说明.

下面引进随机变量 ξ 的特征函数.

定义 若随机变量 ξ 的分布函数为 $F_\xi(x)$, 则称

$$f_\xi(t)=Ee^{it\xi}=\int_{-\infty}^{\infty}e^{itx}dF_\xi(x) \tag{2}$$

为 ξ 的特征函数.

特征函数是一个实变量的复值函数, 由于 $|e^{itx}|=1$, 所以它对一切实数 t 都有定义.

显然特征函数只与分布函数有关, 因此亦称某一分布函数的特征函数.

对于离散型随机变量, 若其分布列为

$$\begin{pmatrix} x_1, & x_2, & \cdots, & x_n, & \cdots \\ p_1, & p_2, & \cdots, & p_n, & \cdots \end{pmatrix}$$

则其特征函数为

$$f(t)=\sum_{j=1}^{\infty}p_je^{itx_j} \tag{3}$$

特别地, 对于整值随机变量, 若其母函数为 $P(s)$, 则 $f(t)=P(e^{it})$.

对于连续型随机变量, 若其分布密度函数为 $p(x)$, 则其特征函数为

$$f(t)=\int_{-\infty}^{\infty}e^{itx}p(x)dx \tag{4}$$

这时, 特征函数是密度函数 $p(x)$ 的富利埃(Fourier)变换.

一般情况下的特征函数可以看作是这种富利埃变换的推广. 富利埃分析是数学中一种非常有力的工具, 它在许多数学分支中都起了重大作用, 以后我们将会看到, 它在概率论中也占有突出的

地位.

下面指出一些重要分布的特征函数.

[例 1] 退化分布 $I(x-c)$ 的特征函数为

$$f(t)=e^{ict} \tag{5}$$

[例 2] 二项分布 $b(k;n,p)$ 的特征函数为

$$f(t)=(pe^{it}+q)^n \tag{6}$$

[例 3] 普阿松分布的特征函数为

$$f(t)=e^{\lambda(e^{it}-1)} \tag{7}$$

[例 4] 正态分布 $N(a,\sigma^2)$ 的特征函数为

$$f(t)=\frac{1}{\sqrt{2\pi}\,\sigma}\int_{-\infty}^{\infty}e^{itx-(x-a)^2/2\sigma^2}dx$$

$$=e^{iat-\sigma^2t^2/2}\cdot\frac{1}{\sqrt{2\pi}}\int_{-\infty-it\sigma}^{\infty-it\sigma}e^{-z^2/2}dz$$

$$=e^{iat-\frac{1}{2}\sigma^2t^2} \tag{8}$$

例 1 的计算是显然的，例 2 及例 3 可以直接计算也可以用母函数的相应公式得出，(8)的计算要用围道积分。

二、性质

下面介绍特征函数的一些基本性质。

性质 1 特征函数 $f(t)$ 有如下性质:

$$f(0)=1 \tag{9}$$

$$|f(t)|\leqslant f(0) \tag{10}$$

$$f(-t)=\overline{f(t)} \tag{11}$$

[证明] $f(0)=\int_{-\infty}^{\infty}1dF(x)=1$

$$|f(t)|\leqslant\int_{-\infty}^{\infty}|e^{itx}|dF(x)=1=f(0)$$

$$f(-t)=\int_{-\infty}^{\infty}e^{-itx}dF(x)$$

$$=\overline{\int_{-\infty}^{\infty} e^{itx}dF(x)}=\overline{f(t)}$$

性质2 特征函数在$(-\infty,\infty)$上一致连续。

〔证明〕 因为

$$|f(t+h)-f(t)|=\left|\int_{-\infty}^{\infty}(e^{i(t+h)x}-e^{itx})dF(x)\right|$$

$$\leqslant\int_{-\infty}^{\infty}|e^{ihx}-1|dF(x)\leqslant 2\int_{|x|\geqslant A}dF(x)+\int_{-A}^{A}|e^{ihx}-1|dF(x)$$

$$=2\int_{|x|\geqslant A}dF(x)+2\int_{-A}^{A}\left|\sin\frac{hx}{2}\right|dF(x)$$

注意上式右边已与t无关；可选足够大的A使$\int_{|x|\geqslant A}dF(x)$任意小，然后选充分小的$|h|$可使第二个积分也任意小，从而证明了结论.

性质3 对于任意的正整数n及任意实数$t_1, t_2, \cdots, t_n$及复数$\lambda_1, \lambda_2, \cdots, \lambda_n$，成立

$$\sum_{k=1}^{n}\sum_{j=1}^{n}f(t_k-t_j)\lambda_k\bar{\lambda}_j\geqslant 0 \tag{12}$$

〔证明〕

$$\sum_{k=1}^{n}\sum_{j=1}^{n}f(t_k-t_j)\lambda_k\bar{\lambda}_j$$

$$=\sum_{k=1}^{n}\sum_{j=1}^{n}\left\{\int_{-\infty}^{\infty}e^{i(t_k-t_j)x}dF(x)\right\}\lambda_k\bar{\lambda}_j$$

$$=\int_{-\infty}^{\infty}\left\{\sum_{k=1}^{n}\sum_{j=1}^{n}e^{i(t_k-t_j)x}\lambda_k\bar{\lambda}_j\right\}dF(x)$$

$$=\int_{-\infty}^{\infty}\left(\sum_{k=1}^{n}e^{it_kx}\lambda_k\right)\left(\sum_{j=1}^{n}e^{-it_jx}\bar{\lambda}_j\right)dF(x)$$

$$=\int_{-\infty}^{\infty}\left|\sum_{k=1}^{n}e^{it_k x}\lambda_k\right|^2 dF(x)\geqslant 0$$

这个性质称为非负定性，以后我们将会看到，这是特征函数最本质的性质之一.

性质 4 两个相互独立的随机变量之和的特征函数等于它们的特征函数之积.

[证明] 设 ξ_1 与 ξ_2 是两个相互独立的随机变量，而 $\eta=\xi_1+\xi_2$，由 ξ_1 与 ξ_2 的独立性不难推得复随机变量 $e^{it\xi_1}$ 与 $e^{it\xi_2}$ 也是独立的，因此

$$Ee^{it\eta}=Ee^{it(\xi_1+\xi_2)}=Ee^{it\xi_1}\cdot Ee^{it\xi_2}$$

性质 4 可推广到 n 个独立随机变量之和的场合.

应当着重指出，正是由于性质 4，才使特征函数在概率论中占有重要地位. 由于这个性质，独立随机变量和的特征函数可以方便地用各个特征函数相乘来求得，而独立和的分布函数要通过褶积这种复杂的运算才能得到，相比之下，用特征函数来处理独立和问题就有力得多. 独立和问题在概率论的古典问题中占有"中心"地位，而这些问题的解决大大有赖于特征函数的引进.

性质 5 设随机变量 ξ 有 n 阶矩存在，则它的特征函数可微分 n 次，且当 $k\leqslant n$ 时：

$$f^{(k)}(0)=i^k E\xi^k \tag{13}$$

[证明]

$$\left|\frac{d^k}{dt^k}(e^{itx})\right|=|i^k x^k e^{itx}|\leqslant|x|^k$$

由于 ξ 的 k 阶矩存在，故 $\int_{-\infty}^{\infty}|x|^k dF(x)<\infty$，因而可作下列积分号下的微分

$$f^{(k)}(t)=\int_{-\infty}^{\infty}\frac{d^k}{dt^k}(e^{itx})dF(x)$$

$$= i^k \int_{-\infty}^{\infty} x^k e^{itx} dF(x)$$

取 $t=0$ 即得(13)。

性质 5 使我们可以方便地求得随机变量的各阶矩。

性质 6　设 $\eta = a\xi + b$，这里 a, b 为常数，则

$$f_\eta(t) = e^{ibt} f_\xi(at) \tag{14}$$

［证明］

$$f_\eta(t) = Ee^{it\eta} = Ee^{it(a\xi+b)}$$
$$= e^{itb} Ee^{ita\xi} = e^{ibt} f(at)$$

三、逆转公式与唯一性定理

现在来证明特征函数与分布函数是相互唯一确定的，由分布函数决定特征函数是显然的，剩下来的是需要证明可由特征函数唯一决定分布函数。

下面定理的证明要用到如下数学分析的引理。

引理　设 $x_1 < x_2$，

$$g(T, x, x_1, x_2) = \frac{1}{\pi} \int_0^T \left[\frac{\sin t(x-x_1)}{t} - \frac{\sin t(x-x_2)}{t} \right] dt \tag{15}$$

则

$$\lim_{T\to\infty} g(T, x, x_1, x_2) = \begin{cases} 0, & x < x_1 \text{ 或 } x > x_2 \\ \frac{1}{2}, & x = x_1 \text{ 或 } x = x_2 \\ 1, & x_1 < x < x_2 \end{cases} \tag{16}$$

［证明］　从数学分析中知道狄里克莱积分

$$D(\alpha) = \frac{1}{\pi} \int_0^{\infty} \frac{\sin \alpha t}{t} dt = \begin{cases} \frac{1}{2}, & \alpha > 0 \\ 0, & \alpha = 0 \\ -\frac{1}{2}, & \alpha < 0 \end{cases} \tag{17}$$

而

$$\lim_{T\to\infty} g(T,x,x_1,x_2)=D(x-x_1)-D(x-x_2)$$

分别考察 x 在区间 $(x_1,\ x_2)$ 的端点及内外时相应狄里克莱积分的值即得(16)。

定理(逆转公式) 设分布函数 $F(x)$ 的特征函数为 $f(t)$，又 x_1,x_2 是 $F(x)$ 的连续点，则

$$F(x_2)-F(x_1)=\lim_{T\to\infty}\frac{1}{2\pi}\int_{-T}^{T}\frac{e^{-itx_1}-e^{-itx_2}}{it}f(t)dt \tag{18}$$

[证明] 不妨设 $x_1<x_2$，由于

$$I_T=\frac{1}{2\pi}\int_{-T}^{T}\frac{e^{-itx_1}-e^{-itx_2}}{it}f(t)dt$$

$$=\frac{1}{2\pi}\int_{-T}^{T}\int_{-\infty}^{\infty}\frac{e^{itx_1}-e^{-itx_2}}{it}e^{itx}dF(x)dt$$

交换上述二次积分顺序得到

$$I_T=\frac{1}{2\pi}\int_{-\infty}^{\infty}\left[\int_{-T}^{T}\frac{e^{-itx_1}-e^{-itx_2}}{it}e^{itx}dt\right]dF(x)$$

$$=\frac{1}{2\pi}\int_{-\infty}^{\infty}\left[\int_{0}^{T}\frac{e^{it(x-x_1)}-e^{-it(x-x_1)}-e^{it(x-x_2)}+e^{-it(x-x_2)}}{it}dt\right]dF(x)$$

$$=\frac{1}{\pi}\int_{-\infty}^{\infty}\left[\int_{0}^{T}\left(\frac{\sin t(x-x_1)}{t}-\frac{\sin t(x-x_2)}{t}\right)dt\right]dF(x)$$

$$=\int_{-\infty}^{\infty}g(T,x,x_1,x_2)dF(x)$$

此处 $g(T,x,x_1,x_2)$ 按(15)定义。由(16)知 $|g(T,x,x_1,x_2)|$ 有界，因此由勒贝格控制收敛定理①并利用引理的结果可得：

$$\lim_{T\to\infty}I_T=\int_{-\infty}^{\infty}\lim_{T\to\infty}g(T,x,x_1,x_2)dF(x)$$

$$=F(x_2)-F(x_1)$$

① 为了避免冗长的分析论证，在证明特征函数的逆转公式及逆极限定理时，我们共四次使用了实变函数论中关于极限号与积分号交换的勒贝格控制收敛定理。

定理(唯一性定理) 分布函数由其特征函数唯一决定.

[证明] 应用逆转公式，在 $F(x)$ 的每一连续点上，当 y 沿 $F(x)$ 的连续点趋于 $-\infty$ 时，有

$$F(x)=\frac{1}{2\pi}\lim_{y\to-\infty}\lim_{T\to\infty}\int_{-T}^{T}\frac{e^{-ity}-e^{-itx}}{it}f(t)dt \tag{19}$$

而分布函数由其连续点上的值唯一决定.

由唯一性定理可知特征函数也完整地描述了随机变量.

特别当 $f(t)$ 是绝对可积函数时，有下列更强的结果.

定理 若 $\int_{-\infty}^{\infty}|f(t)|dt<\infty$，则相应的分布函数 $F(x)$ 的导数存在并连续，而且

$$F'(x)=\frac{1}{2\pi}\int_{-\infty}^{\infty}e^{-itx}f(t)dt \tag{20}$$

[证明] 由逆转公式，若 $x-\delta$ 及 $x+\delta$ 是 $F(x)$ 的连续点，则

$$\begin{aligned}&F(x+\delta)-F(x-\delta)\\&=\lim_{T\to\infty}\frac{1}{\pi}\int_{-T}^{T}\frac{\sin t\delta}{t}e^{-itx}f(t)dt\end{aligned} \tag{21}$$

因此

$$\begin{aligned}&\frac{F(x+\delta)-F(x-\delta)}{2\delta}\\&=\lim_{T\to\infty}\frac{1}{2\pi}\int_{-T}^{T}\frac{\sin t\delta}{t\delta}e^{-itx}f(t)dt\end{aligned}$$

由于 $\left|\frac{\sin t\delta}{t\delta}e^{-itx}f(t)\right|\leqslant|f(t)|$，因此用控制收敛定理知

$$\begin{aligned}F'(x)&=\lim_{\delta\to0}\frac{F(x+\delta)-F(x-\delta)}{2\delta}\\&=\frac{1}{2\pi}\int_{-\infty}^{\infty}e^{-itx}f(t)dt\end{aligned}$$

而 $F'(x)$ 的连续性可再次用控制收敛定理证得.

因此在$f(t)$是绝对可积的条件下，分布密度$p(x)$与特征函数$f(t)$通过富利埃变换来联系.

四、分布函数的再生性

许多重要的分布函数具有一个有趣的性质——再生性. 这个性质用特征函数来研究最为方便. 下面通过几个例子来说明它.

[例 5] 若ξ_1服从$b(k;n,p)$，ξ_2服从$b(k;m,p)$，而且ξ_1与ξ_2独立，则$\eta=\xi_1+\xi_2$服从$b(k;m+n,p)$.

事实上$f_{\xi_1}(t)=(pe^{it}+q)^n$，$f_{\xi_2}(t)=(pe^{it}+q)^m$，由性质 4 知

$$f_\eta(t)=(pe^{it}+q)^{m+n}$$

因此由唯一性定理知η服从$b(k;m+n,p)$.

[例 6] 若ξ_1服从$p(k;\lambda_1)$，ξ_2服从$p(k;\lambda_2)$，而且ξ_1与ξ_2独立，则$\eta=\xi_1+\xi_2$服从$p(k;\lambda_1+\lambda_2)$. 事实上

$$f_{\xi_1}(t)=e^{\lambda_1(e^{it}-1)},\quad f_{\xi_2}(t)=e^{\lambda_2(e^{it}-1)}$$

$$f_\eta(t)=e^{(\lambda_1+\lambda_2)(e^{it}-1)}$$

[例 7] 若ξ_1服从$N(a_1,\sigma_1^2)$，ξ_2服从$N(a_2,\sigma_2^2)$，而且ξ_1与ξ_2独立，则$\eta=\xi_1+\xi_2$服从$N(a_1+a_2,\sigma_1^2+\sigma_2^2)$.

还有不少重要分布也有再生性，我们把它们留给读者作为练习.

还有人研究了这类命题的逆命题——分布函数的分解问题，即若两个独立随机变量之和服从某一分布，问是否能断定这两个随机变量也分别服从这个分布. 已经证明对于正态分布及普阿松分布逆命题的确成立.

五、多元特征函数

若随机向量$(\xi_1,\xi_2,\cdots,\xi_n)$的分布函数为$F(x_1,x_2,\cdots,x_n)$，与随机变量相仿，我们可以定义它的特征函数

$$f(t_1,t_2,\cdots,t_n)=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}e^{i(t_1x_1+\cdots+t_nx_n)}dF(x_1,\cdots,x_n) \quad (22)$$

可以类似于一元的场合，建立起 n 元特征函数的理论，由于方法完全相同，我们只叙述一些有关结论，证明一概从略。

性质 1 $f(t_1, t_2, \cdots, t_n)$在 R^n 中一致连续，而且

$$|f(t_1, t_2, \cdots, t_n)| \leqslant f(0, 0, \cdots, 0) = 1$$

$$f(-t_1, -t_2, \cdots, -t_n) = \overline{f(t_1, t_2, \cdots, t_n)}$$

性质 2 如果 $f(t_1, t_2, \cdots, t_n)$是$(\xi_1, \xi_2, \cdots, \xi_n)$的特征函数，则 $\eta = a_1\xi_1 + a_2\xi_2 + \cdots + a_n\xi_n$ 的特征函数为

$$f_\eta(t) = f(a_1t, a_2t, \cdots, a_nt) \tag{23}$$

性质 3 如果矩 $E\xi_1^{k_1}\cdot\xi_2^{k_2}\cdots\xi_n^{k_n}$ 存在，则

$$E\xi_1^{k_1}\cdot\xi_2^{k_2}\cdots\xi_n^{k_n}$$

$$= i^{-\sum\limits_{j=1}^{n} k_j}\left[\frac{\partial^{k_1+k_2+\cdots+k_n} f(t_1, t_2, \cdots, t_n)}{\partial t_1^{k_1}\partial t_2^{k_2}\cdots\partial t_n^{k_n}}\right]_{t_1=t_2=\cdots=t_n=0} \tag{24}$$

性质 4 若$(\xi_1, \xi_2, \cdots, \xi_n)$的特征函数为 $f(t_1, t_2, \cdots, t_n)$，则 k $(k<n)$维随机向量$(\xi_1, \xi_2, \cdots, \xi_k)$的特征函数为

$$f_{1,2,\cdots,k}(t_1, t_2, \cdots, t_k) = f(t_1, t_2, \cdots, t_k, 0, \cdots, 0) \tag{25}$$

这是前 k 个分量的 k 维边际分布所对应的特征函数．对应于任意 k 个分量 $\xi_{j_1}, \xi_{j_2}, \cdots, \xi_{j_k}$ 的边际分布的特征函数，可以类似得到．

逆转公式 如果 $f(t_1, t_2, \cdots, t_n)$是随机向量$(\xi_1, \xi_2, \cdots, \xi_n)$的特征函数，而 $F(x_1, x_2, \cdots, x_n)$是它的分布函数，则

$$P\{a_k \leqslant \xi_k < b_k,\ \ k = 1, 2, \cdots, n\}$$

$$= \lim_{\substack{T_j\to\infty\\ j=1,\cdots,n}} \frac{1}{(2\pi)^n}\int_{-T_1}^{T_1}\int_{-T_2}^{T_2}\cdots\int_{-T_n}^{T_n}\prod_{k=1}^{n}\frac{e^{-it_ka_k} - e^{-it_kb_k}}{it_k}$$

$$\cdot f(t_1, t_2, \cdots, t_n)dt_1dt_2\cdots dt_n \tag{26}$$

其中 a_k 和 b_k 都是任意实数，但满足唯一的要求：$(\xi_1, \xi_2, \cdots, \xi_n)$落在平行体

$$a_k \leqslant x_k < b_k, \quad k=1,2,\cdots,n$$

的面上的概率等于零.

唯一性定理 分布函数 $F(x_1, x_2, \cdots, x_n)$由其特征函数唯一决定.

有了唯一性定理,可以进一步证明特征函数的如下两个性质,它们表征了独立性.

性质5 若$(\xi_1, \xi_2, \cdots, \xi_n)$的特征函数为$f(t_1, t_2, \cdots, t_n)$,而$\xi_j$的特征函数为$f_{\xi_j}(t)$, $j=1,2,\cdots,n$,则随机变量$\xi_1, \xi_2, \cdots, \xi_n$相互独立的充要条件为

$$f(t_1, t_2, \cdots, t_n)=f_{\xi_1}(t_1)f_{\xi_2}(t_2)\cdots f_{\xi_n}(t_n) \tag{27}$$

性质6 若以$f_1(t_1, \cdots, t_n)$, $f_2(u_1, \cdots, u_m)$及$f(t_1, \cdots, t_n, u_1, \cdots, u_m)$分别记随机向量$(\xi_1, \cdots, \xi_n)$, $(\eta_1, \cdots, \eta_m)$及$(\xi_1, \cdots, \xi_n, \eta_1, \cdots, \eta_m)$的特征函数,则$(\xi_1, \cdots, \xi_n)$与$(\eta_1, \cdots, \eta_m)$独立的充要条件为:对一切实数$t_1, \cdots, t_n$及$u_1, \cdots, u_m$成立

$$\begin{aligned}&f(t_1, \cdots, t_n, u_1, \cdots, u_m)\\&=f_1(t_1, \cdots, t_n)f_2(u_1, \cdots, u_m)\end{aligned} \tag{28}$$

在下一节,我们要用到如下定理,相应于此定理的一维结果,将在下章叙述并证明.

连续性定理 若特征函数列$\{f_k(t_1, t_2, \cdots, t_n)\}$收敛于一个连续函数$f(t_1, t_2, \cdots, t_n)$,则函数$f(t_1, t_2, \cdots, t_n)$是某分布函数所对应的特征函数.

*§5. 多元正态分布

一、密度函数与特征函数

在本节中,我们将讨论多元正态分布的定义与性质,假定读者具有矩阵论的基本知识.因为对二元的场合我们已直接推导过这当中的大部分结果,所以基础知识不足的读者不妨跳过本节.

下面，我们将以黑体的小写字母记列向量

$$\boldsymbol{a}=\begin{pmatrix}a_1\\a_2\\\vdots\\a_n\end{pmatrix},\quad \boldsymbol{\xi}=\begin{pmatrix}\xi_1\\\xi_2\\\vdots\\\xi_n\end{pmatrix}$$

以黑体的大写字母记矩阵，例如

$$\boldsymbol{B}=\begin{pmatrix}b_{11} & b_{12}\cdots b_{1n}\\ b_{21} & b_{22}\cdots b_{2n}\\ \cdots\cdots & \cdots\cdots\cdots\\ b_{n1} & b_{n2}\cdots b_{nn}\end{pmatrix}$$

以 τ 记矩阵（当然也包括向量）的转置，因此 $\boldsymbol{t}^{\tau}\boldsymbol{\xi}=\sum_{i=1}^{n}t_i\xi_i$（其中，$\boldsymbol{t}$ 为列向量）；以 $\boldsymbol{B}^{-1}$ 记 $\boldsymbol{B}$ 的逆阵；以 $|\boldsymbol{B}|$ 记 $\boldsymbol{B}$ 的行列式之值.

在(3.2.13)中，我们已定义了 n 元正态分布的密度函数，采用列向量形式，其表达式为

$$p(\boldsymbol{x})=\frac{1}{(2\pi)^{n/2}|\boldsymbol{B}|^{\frac{1}{2}}}\exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{a})^{\tau}\boldsymbol{B}^{-1}(\boldsymbol{x}-\boldsymbol{a})\right\}\tag{1}$$

其中 $\boldsymbol{B}$ 是 n 阶正定对称矩阵，$\boldsymbol{a}$ 是实值列向量，并简记为 $N(\boldsymbol{a},\boldsymbol{B})$.

事实上，我们还需要证明由(1)定义的函数是 R^n 中的密度函数.

显然

$$p(\boldsymbol{x})>0,\quad \boldsymbol{x}\in R^n\tag{2}$$

因此只须验证下式成立

$$\int_{R^n}p(\boldsymbol{x})d\boldsymbol{x}=1\tag{3}$$

为了证明(3)，我们要用到矩阵论中这么一个结果：若 $\boldsymbol{B}$ 是正定对称阵，则存在非奇异阵 $\boldsymbol{L}$，使

$$B=LL'\tag{4}$$

作线性变换

$$y=L^{-1}(x-a)\tag{5}$$

则逆变换为

$$x=Ly+a\tag{6}$$

变换(6)的雅可比行列式由下式给出

$$|L|=|B|^{\frac{1}{2}}\tag{7}$$

因此

$$\int_{R^n}p(x)dx=\frac{1}{(2\pi)^{n/2}|B|^{\frac{1}{2}}}\int_{R^n}\exp\left\{-\frac{1}{2}y'y\right\}\cdot|B|^{\frac{1}{2}}dy$$

$$=\frac{1}{(2\pi)^{n/2}}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\exp\left\{-\frac{1}{2}\sum_{k=1}^{n}y_k^2\right\}dy_1\cdots dy_n$$

$$=\left[\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{-u^2/2}du\right]^n=1$$

从而证明了(3),所以(1)确实定义了 R^n 上的一个密度函数。

定理1 n 元正态分布(1)的特征函数为

$$f(t)=\exp\left\{ia't-\frac{1}{2}t'Bt\right\}\tag{8}$$

[证明] 按定义

$$f(t)=\int_{R^n}e^{it'x}p(x)dx$$

$$=\frac{1}{(2\pi)^{n/2}|B|^{\frac{1}{2}}}$$

$$\cdot\int_{R^n}e^{it'x}\exp\left\{-\frac{1}{2}(x-a)'B^{-1}(x-a)\right\}dx$$

作变换(6),注意到

$$it'x-\frac{1}{2}(x-a)'B^{-1}(x-a)$$

$$= it'a + it'Ly - \frac{1}{2}y'y$$

记 $s = L't$，则

$$it'x - \frac{1}{2}(x-a)'B^{-1}(x-a)$$

$$= it'a + is'y - \frac{1}{2}y'y$$

$$= ia't + \sum_{k=1}^{n}\left(is_k y_k - \frac{1}{2}y_k^2\right)$$

$$= ia't - \frac{1}{2}\sum_{k=1}^{n}(y_k - is_k)^2 - \frac{1}{2}\sum_{k=1}^{n}s_k^2$$

$$= ia't - \frac{1}{2}\sum_{k=1}^{n}(y_k - is_k)^2 - \frac{1}{2}t'Bt$$

因此

$$f(t) = \frac{e^{ia^\tau t - \frac{1}{2}t^\tau Bt}}{(2\pi)^{n/2}|B|^{\frac{1}{2}}}$$

$$\cdot\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\exp\left\{-\frac{1}{2}\sum_{k=1}^{n}(y_k - is_k)^2\right\}|B|^{\frac{1}{2}}dy_1\cdots dy_n$$

$$= e^{ia^\tau t - \frac{1}{2}t^\tau Bt}$$

在(1)中，假定 B 是正定对称阵，否则该表达式没有意义，因此我们只是对正定对称阵 B 的场合定义了多元正态分布，但是利用特征函数表达式(8)有可能把定义拓广到一般非负定对称阵 B 的场合．

事实上，设 B 是非负定对称阵，令 $B_k = B + \frac{1}{k}I$，这里 I 是 n 阶单位阵，显然 B_k 是正定对称阵，因此

$$f_k(t) = e^{ia^\tau t - \frac{1}{2}t^\tau B_k t}$$

是 n 元正态分布 $N(a, B_k)$ 的特征函数．现在

$$\lim_{k\to\infty} f_k(t) = e^{ia^\tau t - \frac{1}{2}t^\tau Bt} = f(t)$$

而 $f(t)$ 在 R^n 上连续，因此由上节的连续性定理知 $f(t)$ 是 R^n 上某分布函数的特征函数．

这样可以引进如下一般定义．

定义　若 a 是 n 维实向量，B 是 n 阶非负定对称阵，则称以(8)式中的 $f(t)$ 为其特征函数的分布函数为 **n 元正态分布**，并简记为 $N(a, B)$．

按照这个定义，当 B 为正定对称阵时，其密度函数由(1)给出；但是当 $|B|=0$ 时，密度函数无法写出．可以证明，若 B 的秩为 r $(r<n)$，则这时概率分布集中在一个 r 维子空间上，这种正态分布称为**退化正态分布或奇异正态分布**．

在下面讨论中，我们总是假定随机向量 $\xi=(\xi_1, \xi_2, \cdots, \xi_n)^\tau$ 服从 n 元正态分布 $N(a, B)$．

定理 2　ξ 的任一子向量 $(\xi_{k_1}, \xi_{k_2}, \cdots, \xi_{k_m})^\tau$，$(m\leqslant n)$ 也服从正态分布，分布为 $N(\tilde{a}, \tilde{B})$，其中 $\tilde{a}=(a_{k_1}, a_{k_2}, \cdots, a_{k_m})^\tau$，$\tilde{B}$ 为保留 B 的第 $k_1, k_2, \cdots, k_m$ 行及列所得的 m 阶矩阵．

特别地，ξ_j 服从一元正态分布 $N(a_j, b_{jj})$．

［证明］　只须在特征函数(8)中对一切不等于 $k_1, k_2, \cdots, k_m$ 的 l，令 $t_l=0$ 即得 $(\xi_{k_1}, \xi_{k_2}, \cdots, \xi_{k_m})^\tau$ 的特征函数

$$\tilde{f}(\tilde{t}) = e^{i\tilde{a}^\tau \tilde{t} - \frac{1}{2}\tilde{t}^\tau \tilde{B}\tilde{t}}$$

这里 $\tilde{t}=(t_{k_1}, t_{k_2}, \cdots, t_{k_m})^\tau$，这正是 $N(\tilde{a}, \tilde{B})$ 的特征函数．

定理 2 表明，多元正态分布的边际分布还是正态分布．

定理 3　a 及 B 分别是随机向量 ξ 的数学期望及协方差矩阵．即

$$a_j = E\xi_j, \qquad 1\leqslant j\leqslant n \tag{9}$$

$$b_{jk}=E(\xi_j-a_j)(\xi_k-a_k),\quad 1\leqslant j,k\leqslant n \tag{10}$$

〔证明〕 由定理2立即得到(9)，而且知道 $D\xi_j=b_{jj}$ 存在，由柯西-许瓦兹不等式又知协方差存在，因此由(4.24)

$$E\xi_j\xi_k=\frac{1}{i^2}\cdot\frac{\partial^2 f(t_1,\cdots,t_n)}{\partial t_j\partial t_k}\bigg|_{t_1=\cdots=t_n=0}=b_{jk}+a_ja_k$$

故

$$E(\xi_j-a_j)(\xi_k-a_k)=E\xi_j\xi_k-a_ja_k=b_{jk}$$

因此 n 元正态分布由它的前面二阶矩完全确定。

二、独立性

正如在二元场合一样，对 n 元正态分布而言，独立性与不相关性有密切联系。

定理4 $\xi_1,\xi_2,\cdots,\xi_n$ 相互独立的充要条件是它们两两不相关。

〔证明〕 必要性显然成立，下证充分性。

若 $\xi_1,\xi_2,\cdots,\xi_n$ 两两不相关，即对一切 $j\neq k$，

$$r_{jk}=\frac{E(\xi_j-E\xi_j)(\xi_k-E\xi_k)}{\sqrt{D\xi_j}\sqrt{D\xi_k}}=0$$

因此 $b_{jk}=E(\xi_j-E\xi_j)(\xi_k-E\xi_k)=0$，所以

$$f(t_1,\cdots,t_n)=e^{i\sum\limits_{k=1}^{n}a_kt_k-\frac{1}{2}\sum\limits_{k=1}^{n}b_{kk}t_k^2}$$

$$=\prod_{k=1}^{n}e^{ia_kt_k-\frac{1}{2}b_{kk}t_k^2}$$

$$=\prod_{k=1}^{n}f_{\xi_k}(t_k)$$

由上节多元特征函数的性质5可知 $\xi_1,\xi_2,\cdots,\xi_n$ 相互独立。

定理5 若 $\xi=\begin{pmatrix}\xi_1\\ \xi_2\end{pmatrix}$，这里 ξ_1 与 ξ_2 是 ξ 的子向量，记

$$B=\begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix} \tag{11}$$

其中 B_{11} 及 B_{22} 分别是 ξ_1 及 ξ_2 的协方差矩阵，B_{12} 则是由 ξ_1 与 ξ_2 的相应分量的协方差构成的相互协方差矩阵，则 ξ_1 与 ξ_2 独立的充要条件是 $B_{12}=0$.

［证明］ 若 ξ_1 与 ξ_2 独立，则 ξ_1 的任一分量与 ξ_2 的任一分量独立，因此其协方差为 0，从而由它们构成的矩阵 $B_{12}=0$，这就证明了必要性.

下面来证明充分性．由 $b_{ij}=b_{ji}$，因此 $B_{21}=B_{12}'=0$，若 $t=\begin{pmatrix} t_1 \\ t_2 \end{pmatrix}$，这里 t_1 与 ξ_1 有相同维数，t_2 与 ξ_2 有相同维数，则

$$t'Bt=t_1'B_{11}t_1+2t_1'B_{12}t_2+t_2'B_{22}t_2=t_1'B_{11}t_1+t_2'B_{22}t_2$$

因此若记 $a=\begin{pmatrix} a_1 \\ a_2 \end{pmatrix}$，其中 a_1, a_2 分别为 ξ_1 及 ξ_2 的数学期望，则

$$\begin{aligned} f_\xi(t) &= \exp\left\{ ia't-\frac{1}{2}t'Bt \right\} \\ &= \exp\left\{ia_1't_1+ia_2't_2-\frac{1}{2}t_1'B_{11}t_1-\frac{1}{2}t_2'B_{22}t_2\right\} \\ &= \exp\left\{ia_1't_1-\frac{1}{2}t_1'B_{11}t_1\right\}\cdot \exp\left\{ia_2't_2-\frac{1}{2}t_2'B_{22}t_2\right\} \\ &= f_{\xi_1}(t_1)f_{\xi_2}(t_2) \end{aligned}$$

由上节多元特征函数的性质 6 可知 ξ_1 与 ξ_2 独立.

类似地可以证明，若 ξ 的子向量 $\xi_1, \xi_2, \cdots, \xi_k$ 两两独立，则它们也相互独立.

三、线性变换

服从正态分布的随机向量在线性变换下具有许多特殊的性质，这些性质有很大的理论和实用价值，下面只讨论这类性质中最基本的一些.

一般，若 $\xi=(\xi_1, \cdots, \xi_n)'$ 是 n 维随机向量，其数学期望为 $\boldsymbol{a}$，协方差矩阵为 $\boldsymbol{B}$.

考虑 ξ 的分量的线性组合 $\zeta=\sum_{j=1}^{n} l_j \xi_j=\boldsymbol{l}'\xi$，显然

$$E\zeta=\sum_{j=1}^{n} l_j a_j=\boldsymbol{l}'\boldsymbol{a} \tag{12}$$

$$D\zeta=\sum_{j=1}^{n} \sum_{k=1}^{n} l_j l_k b_{jk}=\boldsymbol{l}'\boldsymbol{B}\boldsymbol{l} \tag{13}$$

同样地，若 $\boldsymbol{C}=(c_{jk})$ 是 $m\times n$ 矩阵，而 $\eta=\boldsymbol{C}\xi$ 则

$$E\eta=\boldsymbol{C}\boldsymbol{a} \tag{14}$$

$$D\eta=\boldsymbol{C}\boldsymbol{B}\boldsymbol{C}' \tag{15}$$

这里 $D\eta$ 记 η 的协方差矩阵.

定理6　$\xi=(\xi_1, \cdots, \xi_n)'$ 服从 n 元正态分布 $N(\boldsymbol{a}, \boldsymbol{B})$ 的充要条件是它的任何一个线性组合 $\zeta=\sum_{j=1}^{n} l_j\xi_j$ 服从一元正态分布 $N\left(\sum_{j=1}^{n} l_j a_j, \sum_{j,k=1}^{n} l_j l_k b_{jk}\right)$.

［证明］　必要性：若 ξ 服从 $N(\boldsymbol{a}, \boldsymbol{B})$，则由(8)

$$Ee^{i\boldsymbol{t}'\xi}=\exp\left\{i\boldsymbol{a}'\boldsymbol{t}-\frac{1}{2}\boldsymbol{t}'\boldsymbol{B}\boldsymbol{t}\right\}$$

取 $\boldsymbol{t}=u\boldsymbol{l}$，这里 u 是实数，则

$$Ee^{iu\zeta}=Ee^{iu\boldsymbol{l}'\xi}=\exp\left\{iu\boldsymbol{a}'\boldsymbol{l}-\frac{1}{2}u^2\boldsymbol{l}'\boldsymbol{B}\boldsymbol{l}\right\} \tag{16}$$

对 u 是任意实数上式都成立，这就说明随机变量 ζ 服从 $N(\boldsymbol{l}'\boldsymbol{a}, \boldsymbol{l}'\boldsymbol{B}\boldsymbol{l})$.

充分性：若 $\zeta=\boldsymbol{l}'\xi$ 服从 $N(\boldsymbol{l}'\boldsymbol{a}, \boldsymbol{l}'\boldsymbol{B}\boldsymbol{l})$，仍有(16)，在该式中取 $u=1$，得

$$Ee^{il'\xi}=\exp\left\{ia'l-\frac{1}{2}l'Bl\right\}$$

由于 l 的任意性，这说明 ξ 服从 $N(a, B)$.

利用定理 6，可以通过一维正态变量来研究多维正态变量，在有些场合这提供了很大的方便.

定理 7 若 $\xi=(\xi_1, \cdots, \xi_n)'$ 服从 n 元正态分布 $N(a, B)$，而 C 为任意 $m\times n$ 阵，则 $\eta=C\xi$ 服从 m 元正态分布 $N(Ca, CBC')$.

[证明] 因为对于任意 m 维实值列向量 t,

$$\begin{aligned}f_\eta(t)&=Ee^{it'\eta}=Ee^{it'C\xi}=Ee^{i(C't)'\xi}\\&=\exp\left\{ia'(C't)-\frac{1}{2}(C't)'B(C't)\right\}\\&=\exp\left\{i(Ca)'t-\frac{1}{2}t'(CBC')t\right\}\end{aligned}\tag{17}$$

按定义，η 服从 m 元正态分布 $N(Ca, CBC')$.

定理 7 表明正态变量在线性变换下还是正态变量，这个性质简称为**正态变量的线性变换不变性.**

推论 1 若 ξ 服从 n 元正态分布 $N(a, B)$，则存在一个正交变换 U，使得 $\eta=U\xi$ 是一个具有独立正态分布分量的随机向量，它的数学期望为 Ua，而它的方差分量是 B 的特征值.

[证明] 从矩阵论知道，对实对称矩阵 B，存在正交阵 U，使 $UBU'=D$，其中

$$D=\begin{pmatrix}d_1 & 0 & \cdots & 0\\ 0 & d_2 & \cdots & 0\\ \cdots & \cdots & \cdots & \cdots\\ 0 & 0 & \cdots & d_n\end{pmatrix}\tag{18}$$

这里 $d_1, d_2, \cdots, d_n$ 是 B 的特征值. 若 B 的秩为 r，则有 r 个特征值不为零. 此处的 U 是以特征向量为列构成的正交阵.

把这里的 U 作为定理 7 中的变换矩阵，则利用该定理的结果即得推论 1.

从推论 1 可以看出，若 $\boldsymbol{B}$ 的秩 $r<n$，则正态分布退化到一个 r 维子空间上.

四、条件分布

若 $\boldsymbol{\xi}=\begin{pmatrix}\boldsymbol{\xi}_1\\ \boldsymbol{\xi}_2\end{pmatrix}$服从 n 元正态分布 $N(\boldsymbol{a},\boldsymbol{B})$，这里 $\boldsymbol{\xi}_1$，$\boldsymbol{\xi}_2$ 是它的子向量，$E\boldsymbol{\xi}_1=\boldsymbol{a}_1$，$E\boldsymbol{\xi}_2=\boldsymbol{a}_2$，协方差矩阵 $\boldsymbol{B}$ 仍由(11)表出.

下面求在给定 $\boldsymbol{\xi}_1=\boldsymbol{x}_1$ 的条件下 $\boldsymbol{\xi}_2$ 的分布密度函数. 假定 $|\boldsymbol{B}|\neq 0$，即只讨论非奇异的场合，这时 $|\boldsymbol{B}_{11}|\neq 0$.

首先，我们来找一个线性变换

$$\boldsymbol{\eta}_1=\boldsymbol{\xi}_1 \tag{19}$$

$$\boldsymbol{\eta}_2=\boldsymbol{T}\boldsymbol{\xi}_1+\boldsymbol{\xi}_2 \tag{20}$$

这个线性变换使 $\boldsymbol{\eta}_1=\boldsymbol{\xi}_1$，此外我们还要求它使得 $\boldsymbol{\eta}_1$ 与 $\boldsymbol{\eta}_2$ 独立. 由定理 7 我们知道 $\boldsymbol{\eta}=\begin{pmatrix}\boldsymbol{\eta}_1\\ \boldsymbol{\eta}_2\end{pmatrix}$是联合正态的，因此根据定理 5，为使 $\boldsymbol{\eta}_1$ 与 $\boldsymbol{\eta}_2$ 独立，只须它们的相互协方差矩阵为零阵. 但是

$$\begin{aligned}&E(\boldsymbol{\eta}_1-E\boldsymbol{\eta}_1)(\boldsymbol{\eta}_2-E\boldsymbol{\eta}_2)'\\ &=E(\boldsymbol{\xi}_1-E\boldsymbol{\xi}_1)(\boldsymbol{T}\boldsymbol{\xi}_1+\boldsymbol{\xi}_2-\boldsymbol{T}E\boldsymbol{\xi}_1+E\boldsymbol{\xi}_2)'\\ &=\boldsymbol{B}_{11}\boldsymbol{T}'+\boldsymbol{B}_{12}\end{aligned} \tag{21}$$

因此应取

$$\boldsymbol{T}=-\boldsymbol{B}_{21}\boldsymbol{B}_{11}^{-1} \tag{22}$$

故

$$\boldsymbol{\eta}_2=-\boldsymbol{B}_{21}\boldsymbol{B}_{11}^{-1}\boldsymbol{\xi}_1+\boldsymbol{\xi}_2 \tag{23}$$

所求的线性变换为

$$\begin{pmatrix}\boldsymbol{\eta}_1\\ \boldsymbol{\eta}_2\end{pmatrix}=\begin{pmatrix}\boldsymbol{I} & \boldsymbol{0}\\ -\boldsymbol{B}_{21}\boldsymbol{B}_{11}^{-1} & \boldsymbol{I}\end{pmatrix}\begin{pmatrix}\boldsymbol{\xi}_1\\ \boldsymbol{\xi}_2\end{pmatrix} \tag{24}$$

为求 $\boldsymbol{\eta}_1$ 及 $\boldsymbol{\eta}_2$ 的分布密度函数，先计算

$$E\boldsymbol{\eta}_1=E\boldsymbol{\xi}_1=\boldsymbol{a}_1,\quad D\boldsymbol{\eta}_1=D\boldsymbol{\xi}_1=\boldsymbol{B}_{11} \tag{25}$$

$$E\eta_2 = -B_{21}B_{11}^{-1}E\xi_1 + E\xi_2 = a_2 - B_{21}B_{11}^{-1}a_1 \tag{26}$$

$$\begin{aligned} D\eta_2 &= E(\eta_2 - E\eta_2)(\eta_2 - E\eta_2)^{\tau} \\ &= E[(\xi_2 - a_2) - B_{21}B_{11}^{-1}(\xi_1 - a_1)] \\ &\quad \cdot [(\xi_2 - a_2) - B_{21}B_{11}^{-1}(\xi_1 - a_1)]^{\tau} \\ &= B_{22} - B_{21}B_{11}^{-1}B_{12} - B_{21}B_{11}^{-1}B_{12} + B_{21}B_{11}^{-1}B_{11}B_{11}^{-1}B_{12} \\ &= B_{22} - B_{21}B_{11}^{-1}B_{12} \end{aligned} \tag{27}$$

正如上述，η_1 与 η_2 是独立的，又变换(24)的雅可比行列式等于 1，所以有

$$p_{\xi}(x_1, x_2) = p_{\eta}(y_1, y_2) = p_{\eta_1}(y_1)p_{\eta_2}(y_2) \tag{28}$$

这里 $y_1 = x_1, y_2 = -B_{21}B_{11}^{-1}x_1 + x_2$，显然 $p_{\xi_1}(x_1) = p_{\eta_1}(y_1)$，因此在给定 $\xi_1 = x_1$ 的条件下，ξ_2 的密度函数

$$\begin{aligned} p(x_2 \mid \xi_1 = x_1) &= \frac{p_{\xi}(x_1, x_2)}{p_{\xi_1}(x_1)} = \frac{p_{\eta_1}(y_1)p_{\eta_2}(y_2)}{p_{\eta_1}(y_1)} \\ &= p_{\eta_2}(y_2) = p_{\eta_2}(x_2 - B_{21}B_{11}^{-1}x_1) \end{aligned}$$

因为 η_2 服从 $N(a_2 - B_{21}B_{11}^{-1}a_1,\ B_{22} - B_{21}B_{11}^{-1}B_{12})$，所以 ξ_2 对于 $\xi_1 = x_1$ 的条件分布是正态分布 $N(a_2 + B_{21}B_{11}^{-1}(x_1 - a_1),\ B_{22} - B_{21}B_{11}^{-1}B_{12})$，我们把这个结果总结为下列定理.

定理 8 若 $\xi = \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix}$ 服从 n 元正态分布 $N(a, B)$，$E\xi_1 = a_1$，$E\xi_2 = a_2$，B 表示成(11)，则在给定 $\xi_1 = x_1$ 下，ξ_2 的条件分布还是正态分布，其条件数学期望

$$E(\xi_2 \mid \xi_1 = x_1) = a_2 + B_{21}B_{11}^{-1}(x_1 - a_1) \tag{29}$$

其条件方差

$$B_{22\cdot1} = B_{22} - B_{21}B_{11}^{-1}B_{12} \tag{30}$$

这里 $E(\xi_2 \mid \xi_1 = x_1)$ 称为 ξ_2 关于 ξ_1 的回归，注意到它是 x_1 的线性函数，又条件方差 $B_{22\cdot1}$ 与 x_1 无关.

我们已对二元正态分布的场合进行过类似的讨论，这类问题

将在数理统计部分继续深入研究.

第四章小结

本章主要讨论了随机变量(或分布函数)的数字特征与特征函数，它们都是概率分布的某种表征. 这些讨论不但深化了对随机变量的认识,同时也为以后的研究作了必要的准备.

数字特征是描述随机变量特征的有效工具，它虽然不象分布函数那样完整地描述了随机变量，但是却具有很多优点：它较集中地反映了随机变量变化的一些平均特征（事实上数字特征多是某种平均值);其次,大部分最重要的分布函数都由一、两个数字特征完全确定,而数字特征较易求得;最后，最常用也是最重要的几个数字特征——数学期望,方差,相关系数——都有明确的概率意义，同时又具有良好的性质. 所有这些都决定了数字特征在概率论与数理统计中的重要地位.

特征函数与分布函数一一对应，它虽然不象分布函数那样有直观的概率意义,但却有好得多的分析性质,因此它是解决某些分布问题的有力工具，特别在处理独立随机变量和的分布问题上占有极重要的地位. 关于特征函数与分布函数关系的另一些方面，将在下章中进一步讨论.

数字特征与特征函数都可以看作是分布函数的某种变换，而母函数则是概率论中引进的第一个变换，它在概率论发展史上有很大作用. 到了现在，母函数以它的简明性以及在某些场合的有力应用而保持了它一定的地位.

我们用特征函数作为工具研究了多元正态分布，建立了它的许多重要性质，这些良好性质从一个方面决定了正态分布在概率论与数理统计中的重要地位，正态分布重要性的另一方面——常见性——将在下章中得到深入的讨论.

最近几十年来，信息，信息的传输、变换与处理的概念正在渗入各种各样的学科，大大加深了人类对客观世界的认识. 信息的概念与随机试验、概率分布等概率论概念密切相关，作为不肯定性度量的熵可以看作是概率空间的某种数字特征，因此我们对这些重要概念作了初步的介绍.

本章的作用是承上启下. 本章引进的两个基本概念——数字特征与特征函数——将在极限定理的研究中起重要作用.

习　题

1. 设随机变量 ξ 只取非负整数值，其概率为 $P\{\xi=k\}=\frac{a^k}{(1+a)^{k+1}}$，$a>0$ 是常数，试求 $E\xi$ 及 $D\xi$.

2. 若事件 A 在第 i 次试验中出现的概率为 p_i，设 μ 是事件 A 在起初 n 次独立试验中的出现次数，试求 $E\mu$ 及 $D\mu$.

3. 设 μ 是事件 A 在 n 次独立试验中的出现次数，在每次试验中 $P(A)=p$，再设随机变量 η 视 μ 取偶数或奇数而取数值 0 及 1，试求 $E\eta$ 及 $D\eta$.

4. 袋中有 k 号的球 k 只，$k=1, 2, \cdots, n$，从中摸出一球，求所得号码的数学期望.

5. 随机变量 μ 取非负整数值 $n\geqslant 0$ 的概率为 $p_n=A\frac{B^n}{n!}$，已知 $E\mu=a$，试决定 A 与 B.

6. 一袋中含有 a 只白球，b 只黑球，从中摸出 c 只($c\leqslant a+b$)，求摸出白球数 μ 的数学期望.

7. 袋中有 n 张卡片，记号码 $1, 2, \cdots, n$，从中有放回地抽出 k 张卡片来，求所得号码之和 μ 的数学期望及方差.

8. 在上题中，若采用不放回摸卡片，求所得号码之和的数学期望与方差.

9. 试证：若取非负整数值的随机变量 ξ 的数学期望存在，则

$$E\xi=\sum_{k=1}^{\infty}P\{\xi\geqslant k\}$$

10. 某城市共有 N 辆汽车，车牌号从 1 到 N，若随机地记下 n 辆车的车

牌号，其最大号码为 ξ，求 $E\xi$.

11. 若随机变量 ξ 服从**拉普拉斯分布**，其密度函数为

$$p(x)=\frac{1}{2\lambda}e^{-|x-\mu|/\lambda},\quad -\infty<x<\infty,\quad \lambda>0$$

试求 $E\xi$ 及 $D\xi$.

12. 若分子的速度的分布密度函数由马克斯威尔分布律给出

$$p(x)=\begin{cases}\frac{4x^2}{\alpha^3\sqrt{\pi}}e^{-x^2/\alpha^2}, & x>0\\ 0, & x\leqslant 0\end{cases}$$

其中 $\alpha>0$ 是常数，试求分子的平均速度和平均动能（假定分子的质量等于 m）.

13. 若 ξ_1,ξ_2 相互独立，均服从 $N(a,\sigma^2)$，试证：

$$E\max(\xi_1,\xi_2)=a+\frac{\sigma}{\sqrt{\pi}}$$

14. 若随机变量 ξ 的分布函数为 $F(x)$，试证：

$$E\xi=\int_0^{\infty}[1-F(x)]dx-\int_{-\infty}^{0}F(x)dx$$

15. 若 $\xi_1,\xi_2,\cdots,\xi_n$ 为正的独立随机变量，服从相同分布，密度函数为 $f(x)$，试证

$$E\left(\frac{\xi_1+\xi_2+\cdots+\xi_k}{\xi_1+\xi_2+\cdots+\xi_n}\right)=\frac{k}{n}$$

16. 袋中装有 N 只球，但其中白球数为随机变量，只知其数学期望为 n，试求从该袋中摸一球得到白球的概率.

17. 甲袋中装有 a 只白球 b 只黑球，乙袋中装有 α 只白球 β 只黑球，现从甲袋中摸出 $c(c\leqslant a+b)$ 只球放入乙袋中，求从乙袋中再摸一球而为白球的概率.

18. 袋中有 a 只白球 b 只黑球，每次摸出一球后总是放入一只白球，这样进行了 n 次之后，再从袋中摸一只球，求它是白球的概率.

19. 甲袋中有 a 只白球 b 只黑球，乙袋中有 c 只白球 d 只黑球，从两袋中各摸出一球，并交换放入另一袋中，这样做了 n 次之后，再从甲袋中摸出一球，求这球是白球的概率.

20. 现有 n 个袋子，各装有 a 只白球 b 只黑球，先从第一个袋子中摸出

一球，记下颜色后就把它放入第二个袋子中，再从第二个袋子中摸出一球，记下颜色后就把它放入第三个袋子中，照这样办法依次摸下去，最后从第 n 个袋子中摸出一球并记下颜色，若在这 n 次摸球中所摸得的白球的总数为 S_n，试求 ES_n.

21. 在物理实验中，为测量某物体的重量，通常要重复测量多次，最后再把测量记录的平均值作为该物体的重量，试说明这样做的道理.

22. 若 $\xi_1, \xi_2, \cdots, \xi_n$ 是独立随机变量，$D\xi_i=\sigma_i^2$，试找"权"$a_1, a_2, \cdots, a_n$ $\left(\text{它们满足}\sum_{i=1}^n a_i=1\right)$，使 $\sum_{i=1}^n a_i\xi_i$ 的方差最小.

23. 设随机变量 $\xi_1, \xi_2, \cdots, \xi_{m+n}$ $(n>m)$ 是独立的，有相同的分布并且有有限的方差，试求 $S=\xi_1+\cdots+\xi_n$ 与 $T=\xi_{m+1}+\xi_{m+2}+\cdots+\xi_{m+n}$ 两和之间的相关系数.

24. 若 ξ 的密度函数是偶函数，且 $E\xi^2<\infty$，试证 $|\xi|$ 与 ξ 不相关，但它们不相互独立.

25. 若 (ξ, η) 的密度函数为

$$p(x,y)=\begin{cases}\frac{1}{\pi}, & x^2+y^2\leq 1\\ 0, & x^2+y^2>1\end{cases}$$ 试验证：ξ 与 η 不相关，但它们不独立.

26. 若 ξ 与 η 都是只能取两个值的随机变量，试证如果它们不相关，则独立.

27. 若 $U=aX+b$, $V=cY+d$，试证 U, V 的相关系数等于 X, Y 的相关系数.

28. 若 ξ_1, ξ_2, ξ_3 是三个随机变量，试讨论(1) ξ_1, ξ_2, ξ_3 两两不相关；(2) $D(\xi_1+\xi_2+\xi_3)=D\xi_1+D\xi_2+D\xi_3$；(3) $E\xi_1\xi_2\xi_3=E\xi_1\cdot E\xi_2\cdot E\xi_3$ 之间的关系.

29. 若 ξ, η 服从二元正态分布，$E\xi=a$, $D\xi=1$, $E\eta=b$, $D\eta=1$，证明：ξ 与 η 的相关系数 $r=\cos q\pi$，其中 $q=P\{(\xi-a)(\eta-b)<0\}$.

30. 设 (ξ, η) 服从二元正态分布，$E\xi=E\eta=0$, $D\xi=D\eta=1$, $r_{\xi\eta}=r$，试证

$$E\max(\xi,\eta)=\sqrt{\frac{1-r}{\pi}}$$

31. 设 ξ, η 独立，均具有相同分布 $N(a,\sigma^2)$，试求 $p\xi+q\eta$ 与 $u\xi+v\eta$ 的相关系数.

32. Pareto 分布的密度函数为

$$p(x)=\begin{cases} rA^r\frac{1}{x^{r+1}}, & x\geq A \\ 0, & x<A \end{cases}$$ 这里 $r>0, A>0$. 试指出这分布具有 p 阶矩，当且仅当 $p<r$.

33. 若 ξ 的密度函数为 $p(x)=\begin{cases} \frac{1}{2|x|(\log|x|)^2}, & |x|>e \\ 0, & 其它 \end{cases}$ 试证对于任何 $a>0, E|\xi|^a=\infty$.

34. 若 ξ 服从 $N(a,\sigma^2)$，试求 $E|\xi-a|^k$.

35. 记 $a_k=E|\xi|^k$，若 $a_n<\infty$，试证 $\sqrt[k]{a_k}\leq\sqrt[k+1]{a_{k+1}}$，$k=1, 2, \cdots, n-1$.

36. 甲袋中装有 5 只白球，7 只黑球，3 只红球，乙袋中装有 4 只白球，4 只黑球，7 只红球，试问从哪一个袋中取出一只球有较大不肯定性?

37. 试求几何分布的熵.

38. 试求二项分布的熵.

39. 若以 α 及 β 分别记二进制信道的输入及输出，已知 $P\{\alpha=1\}=p$, $P\{\alpha=0\}=1-p$, $P\{\beta=1|\alpha=1\}=q$, $P\{\beta=0|\alpha=1\}=1-q$, $P\{\beta=1|\alpha=0\}=r$, $P\{\beta=0|\alpha=0\}=1-r$，试求输出中含有输入的信息量.

40. 在 12 只金属球中，混有一只假球，并且不知道它是比真球轻还是重，用没有砝码的天平来称这些球，试问至少需要称多少次才能查出这个假球并确定它是比真球轻或重.

41. 试用母函数法求巴斯卡分布的数学期望及方差.

42. 设 ξ 是一个母函数为 $P(s)$ 的随机变量，试求下列各概率所对应的母函数:

(1) $P\{\xi>n\}$;　　　　(2) $P\{\xi=2n\}$.

43. 在贝努里试验中，若试验次数 ν 是随机变量，试证成功的次数与失败的次数这两个随机变量独立的充要条件是 ν 服从普阿松分布.

44. 设 $\{\xi_k\}$ 是一串独立的整值随机变量序列，具有相同概率分布，考虑和 $\eta=\xi_1+\xi_2+\cdots+\xi_\nu$，其中 ν 是随机变量，它与 $\{\xi_k\}$ 相互独立，试用(1)母函数法；(2)直接计算证明

$$E\eta=E\nu\cdot E\xi_k,\quad D\eta=E\nu\cdot D\xi_k+D\nu\cdot(E\xi_k)^2$$

45. 某公共汽车站在[0, t]中来到的乘客批数 μ 服从参数为 λt 的普阿

松分布，而每批来到的乘客数是随机变量，来 n 个的概率为 p_n，$n=0,1,2,\cdots$，试求 $[0,t]$ 中来到乘客数 η 的母函数及数学期望.

46. 试用母函数法证明二项分布及普阿松分布的再生性.

47. 若分布函数 $F(x)=1-F(-x+0)$ 成立，则称它是**对称的**；试证分布函数对称的充要条件是它的特征函数是实的偶函数.

48. 试求 $[0,1]$ 均匀分布的特征函数.

49. 一般柯西分布的密度函数为 $p(x)=\frac{1}{\pi}\cdot\frac{\lambda}{\lambda^2+(x-\mu)^2}$，$\lambda>0$，试证它的特征函数为 $e^{i\mu t-\lambda|t|}$，利用这个结果证明柯西分布的再生性.

50. 若随机变量 ξ 服从柯西分布，$\mu=0$，$\lambda=1$，而 $\eta=\xi$，试证关于特征函数成立着

$$f_{\xi+\eta}(t)=f_\xi(t)\cdot f_\eta(t)$$

但是 ξ 与 η 并不独立.

51. 试求指数分布与 Γ-分布的特征函数，并证明对于具有相同 λ 值的 Γ-分布，关于参数 r 有再生性.

52. 若 $\xi_1,\xi_2,\cdots,\xi_n$ 相互独立且服从相同分布 $N(0,1)$，试证 $\eta=\sum_{i=1}^{n}\xi_i^2$ 服从参数为 n 的 χ^2-分布，并说明 χ^2-分布也有再生性.

53. 求证：对于任何实值特征函数 $f(t)$，以下两个不等式成立

$$1-f(2t)\leqslant 4(1-f(t))$$

$$1+f(2t)\geqslant 2(f(t))^2$$

54. 求证，如果 $f(t)$ 是相应于分布函数 $F(x)$ 的特征函数，则对于任何 x 值恒成立

$$\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}f(t)e^{-itx}dt=F(x+0)-F(x-0)$$

55. 随机变量 ξ 的特征函数为 $f(t)$，且它的 n 阶矩存在，令

$$\chi_k=\frac{1}{i^k}\left[\frac{d^k}{dt^k}\log f(t)\right]_{t=0},\quad k\leqslant n$$

称 χ_k 为随机变量 ξ 的 k **阶半不变量**. 试证 $\eta=\xi+b$（b 是常数）的 $k(k>1)$ 阶半不变量等于 χ_k.

56. 试求出半不变量与原点矩之间的关系式.

57. 若(ξ,η)服从二元正态分布，试求$\zeta=\xi+\eta$的分布密度函数。

58. 设$\xi_1,\xi_2,\cdots,\xi_n$相互独立，具有相同分布$N(a,\sigma^2)$，试求$\boldsymbol{\xi}=\begin{pmatrix}\xi_1\\ \xi_2\\ \vdots\\ \xi_n\end{pmatrix}$的分布，并写出它的数学期望及协方差阵。再求$\bar{\xi}=\frac{1}{n}\sum_{i=1}^{n}\xi_i$的分布密度。

59. 若$\boldsymbol{\xi}$服从二元正态分布$N(\mathbf{0},\boldsymbol{\Sigma})$，其中

$$\boldsymbol{\Sigma}=\begin{pmatrix}4 & 2\\ 2 & 1\end{pmatrix}$$

试找出矩阵$\boldsymbol{A}$，使$\boldsymbol{\xi}=\boldsymbol{A}\boldsymbol{\eta}$，且要求$\boldsymbol{\eta}$服从非退化的正态分布，并求$\boldsymbol{\eta}$的密度函数。

60. 证明：在正交变换下，多元正态分布的独立、同方差性不变。

第五章　极限定理

§1. 贝努里试验场合的极限定理

一、问题的提出

在第一章中我们已经指出：人们在长期实践中发现，虽然个别随机事件在某次试验中可以出现也可以不出现，但是在大量重复试验中却呈现出明显的规律性，即一个随机事件出现的频率在某个固定数的附近摆动，这就是所谓"频率稳定性"，对于这点，至今为止，我们没有给予理论上的说明．

数学上怎样来描述在一定条件下的大量重复试验呢？我们在第二章已经建立了贝努里试验这一概率模型，并指出它可以作为在一定条件下的重复试验的数学模型．在贝努里试验中，各次试验是相互独立的，在每次试验中，我们所关心的事件A出现的概率$P(A)=p$保持不变，这些特征可以看作是从数学角度把"在一定条件下"、"重复试验"等等用语涵义加以明确化．

在贝努里试验中，若以μ_n记n次试验中A出现的次数，则$\frac{\mu_n}{n}$便是在这n次试验中事件A出现的频率，所谓频率稳定性无非是指当试验次数n增大时，频率$\frac{\mu_n}{n}$接近于某个固定的常数．

这个固定的常数就是事件A在一次试验中发生的概率．由此可见，讨论频率$\frac{\mu_n}{n}$的极限行为是理解概率论中最基本的概念——概率，所不可缺少的．正是这个缘故，在概率论的发展史上，极限定理的研究一直占重要地位，而它的发源地就是贝努里试验这个

概型.

从前几章的讨论中我们知道，μ_n 是随机变量，它服从二项分布

$$P\{\mu_n=k\}=\binom{n}{k}p^k q^{n-k},\quad k=0,1,2,\cdots,n$$

其数学期望 $E\mu_n=np$，方差 $D\mu_n=npq$. 这在一定程度上帮助我们进一步了解了频率 $\frac{\mu_n}{n}$ 的性质. 但是我们更需要知道的是 n 很大时 μ_n 或 $\frac{\mu_n}{n}$ 的性质.

显然，当 n 很大时，μ_n 一般也很大，所以直接研究 μ_n 不很恰当，还是研究频率 $\frac{\mu_n}{n}$ 为宜. 因为 $E\frac{\mu_n}{n}=p$，$D\frac{\mu_n}{n}=\frac{pq}{n}$，所以当 $n\to\infty$ 时，频率的数学期望保持不变，而方差则趋于 0. 我们知道方差为 0 的随机变量是常数，于是我们自然预期频率将趋于常数 p（即事件 A 发生的概率）. 但是频率 $\frac{\mu_n}{n}$ 是随机变量，关于它的极限又将如何提法呢？

一种提法是：当 n 足够大时，频率 $\frac{\mu_n}{n}$ 与概率 p 有较大偏差的概率很小. 用数学语言来讲，就是要证明：对于任意 $\varepsilon>0$，

$$\lim_{n\to\infty}P\left\{\left|\frac{\mu_n}{n}-p\right|\geqslant\varepsilon\right\}=0\tag{1}$$

或者它的等价的式子成立，即

$$\lim_{n\to\infty}P\left\{\left|\frac{\mu_n}{n}-p\right|<\varepsilon\right\}=1\tag{2}$$

历史上，贝努里第一个研究了这种类型的极限定理，在 1713 年发表的论文中（这是概率论的第一篇论文[1]），他建立了(1)，这是一大类概率论极限定理——大数定律中的第一个.

关于频率接近于概率还有其他提法，譬如波雷尔建立了

$$P\left\{\lim_{n\to\infty}\frac{\mu_n}{n}=p\right\}=1 \tag{3}$$

从而开创了另一种形式的极限定理——强大数定律的研究。

本节只讨论贝努里大数定律，关于强大数定律的讨论将在 § 4 中进行。

为了研究 μ_n 的极限行为，可以讨论它的分布 $P\{\mu_n<x\}$ 的变化情况．但是由于 $E\mu_n=np$, $D\mu_n=npq$，因此对于固定的 x 来考虑 $P\{\mu_n<x\}$ 的极限不会有多大意义，因为它将趋于 0，所以通常改为研究"标准化"的随机变量

$$\zeta_n=\frac{\mu_n-np}{\sqrt{npq}} \tag{4}$$

的分布函数

$$P\{\zeta_n<x\}$$

的极限行为，由 ζ_n 的分布函数不难求得 μ_n 的分布函数．

关于上述分布，已证明它的极限分布是正态分布 $N(0,1)$，即

$$\lim_{n\to\infty}P\{\zeta_n<x\}=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-t^2/2}dt \tag{5}$$

这个结果最早由法国数学家德莫哇佛(De Moivre)建立，他对 $p=\frac{1}{2}$ 证得了上述结果．后来，拉普拉斯推广到 $0<p<1$ 的一般场合，那是另一类概率论极限定理——中心极限定理中的第一个．

下面将会看到极限定理 (5) 的研究直接联系到大 n 场合的二项分布的计算，这就顺带解决了第二章遗留下来的一些计算问题．

形如(5)的收敛于正态分布的极限定理的研究，在长达两个世纪的时期内成了概率论研究的中心课题，因此得到了中心极限定理的名称．

经过长期的研究，人们认识到，μ_n 具有性质(1)及(5)是由于它是独立随机变量之和，事实上，若令

$$\xi_i=\begin{cases}1, & \text{第 } i \text{ 次试验出现 } A\\ 0, & \text{第 } i \text{ 次试验不出现 } A\end{cases} \tag{6}$$

则

$$\mu_n=\xi_1+\xi_2+\cdots+\xi_n \tag{7}$$

这里 $\xi_1, \xi_2, \cdots, \xi_n$ 是相互独立的．以后我们将会看到，对于一般的随机变量 ξ_i, $i=1, 2, \cdots, n$, 也可以研究它们的和的极限定理，并且在一定条件下，这个和也具有类似于 μ_n 的性质．关于这些问题的研究就成了本章的主要内容．

为了叙述方便起见，我们引进如下定义．

定义 若 $\xi_1, \xi_2, \cdots, \xi_n, \cdots$ 是随机变量序列，令

$$\eta_n=\frac{\xi_1+\xi_2+\cdots+\xi_n}{n} \tag{8}$$

如果存在这样的一个常数序列 $a_1, a_2, \cdots, a_n, \cdots$, 对任意的 $\varepsilon>0$, 恒有

$$\lim_{n\to\infty} P\{|\eta_n-a_n|<\varepsilon\}=1 \tag{9}$$

则称序列 $\{\xi_n\}$ 服从**大数定律(或大数法则)**.

在以后的讨论中，我们几乎总是假定 $\xi_1, \xi_2, \cdots, \xi_n, \cdots$ 是独立随机变量序列．显然，贝努里大数定律是一般大数定律的一种特殊场合．

关于中心极限定理，我们总是对独立随机变量序列 $\xi_1, \xi_2, \cdots, \xi_n, \cdots$ 进行讨论，假定 $E\xi_i$ 及 $D\xi_i$ 存在，令

$$\zeta_n=\frac{\sum_{i=1}^{n}\xi_i-\sum_{i=1}^{n}E\xi_i}{\sqrt{\sum_{i=1}^{n}D\xi_i}} \tag{10}$$

我们的目的是寻找使

$$\lim_{n\to\infty} P\{\zeta_n<x\}=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-t^2/2}dt \tag{11}$$

成立的条件.

一般，若独立随机变量序列 $\xi_1, \xi_2, \cdots, \xi_n, \cdots$ 的标准化和 ζ_n 使(11)成立，则我们称 $\{\xi_i\}$ 服从**中心极限定理**.

显然，德莫哇佛-拉普拉斯极限定理是中心极限定理的特例.

二、贝努里大数定律

我们将证明一个比贝努里大数定律更强的命题，为此需要一个很重要的不等式.

1. 车贝晓夫(Чебышев)不等式

对于任何具有有限方差的随机变量 ξ，都有

$$P\{|\xi - E\xi| \geqslant \varepsilon\} \leqslant \frac{D\xi}{\varepsilon^2} \tag{12}$$

其中 ε 是任一正数.

［证明］ 若 $F(x)$ 是 ξ 的分布函数，则显然有

$$\begin{aligned} P\{|\xi - E\xi| \geqslant \varepsilon\} &= \int_{|x - E\xi| \geqslant \varepsilon} dF(x) \\ &\leqslant \int_{|x - E\xi| \geqslant \varepsilon} \frac{(x - E\xi)^2}{\varepsilon^2} dF(x) \\ &\leqslant \frac{1}{\varepsilon^2} \int_{-\infty}^{\infty} (x - E\xi)^2 dF(x) = \frac{D\xi}{\varepsilon^2} \end{aligned} \tag{13}$$

这就证得了不等式(12). 有时把(12)改写成

$$P\{|\xi - E\xi| < \varepsilon\} \geqslant 1 - \frac{D\xi}{\varepsilon^2} \tag{14}$$

或

$$P\left\{\left|\frac{\xi - E\xi}{\sqrt{D\xi}}\right| \geqslant \delta\right\} \leqslant \frac{1}{\delta^2} \tag{15}$$

车贝晓夫不等式利用随机变量 ξ 的数学期望 $E\xi$ 及方差 $D\xi = \sigma^2$ 对 ξ 的概率分布进行估计. 例如(15)断言不管 ξ 的分布是什么，ξ 落在 $(E\xi - \sigma\delta, E\xi + \sigma\delta)$ 中的概率不小于 $1 - \frac{1}{\delta^2}$. 因为车贝

晓夫不等式只利用数学期望及方差就描述了随机变量的变化情况，因此它在理论研究及实际应用中很有价值.

从车贝晓夫不等式还可以看出，当方差愈小时，事件$\{|\xi-E\xi|\geqslant\varepsilon\}$的概率也愈小，从这里可以看出方差是描述随机变量与其期望值离散程度的一个量，这与我们以前的理解完全一致.

2. 大数定律的证明

车贝晓夫大数定律 设$\xi_1,\xi_2,\cdots,\xi_n,\cdots$是由两两不相关的随机变量所构成的序列，每一随机变量都有有限的方差，并且它们有公共上界

$$D\xi_1\leqslant C,\quad D\xi_2\leqslant C,\quad\cdots,\quad D\xi_n\leqslant C,\quad\cdots$$

则对任意的$\varepsilon>0$，皆有

$$\lim_{n\to\infty}P\left\{\left|\frac{1}{n}\sum_{k=1}^{n}\xi_k-\frac{1}{n}\sum_{k=1}^{n}E\xi_k\right|<\varepsilon\right\}=1 \tag{16}$$

[证明] 因为$\{\xi_k\}$两两不相关，故

$$D\left(\frac{1}{n}\sum_{k=1}^{n}\xi_k\right)=\frac{1}{n^2}\sum_{k=1}^{n}D\xi_k\leqslant\frac{C}{n}$$

再由车贝晓夫不等式得到

$$P\left\{\left|\frac{1}{n}\sum_{k=1}^{n}\xi_k-\frac{1}{n}\sum_{k=1}^{n}E\xi_k\right|<\varepsilon\right\}\geqslant1-\frac{D\left(\frac{1}{n}\sum_{k=1}^{n}\xi_k\right)}{\varepsilon^2}\geqslant1-\frac{C}{n\varepsilon^2}$$

所以

$$1\geqslant P\left\{\left|\frac{1}{n}\sum_{k=1}^{n}\xi_k-\frac{1}{n}\sum_{k=1}^{n}E\xi_k\right|<\varepsilon\right\}\geqslant1-\frac{C}{n\varepsilon^2}$$

于是，当$n\to\infty$时有(16)，因此定理得证.

这个结果在1866年被俄国数学家车贝晓夫所证明，它是关于大数定律的一个相当普遍的结论，许多大数定律的古典结果是它

的特例;此外,证明这个定律所用的方法也很有创造性,在这基础上发展起来的一系列不等式是研究各种极限定理的有力工具.

贝努里大数定律 设 μ_n 是 n 次贝努里试验中事件 A 出现的次数,而 p 是事件 A 在每次试验中出现的概率,则对任意 $\varepsilon>0$,都有

$$\lim_{n\to\infty} P\left\{\left|\frac{\mu_n}{n}-p\right|<\varepsilon\right\}=1 \tag{17}$$

[证明] 定义随机变量 ξ_i 如(6),则

$$E\xi_i=p, \qquad D\xi_i=pq\leqslant\frac{1}{4}$$

而

$$\frac{1}{n}\sum_{k=1}^{n}\xi_k-\frac{1}{n}\sum_{k=1}^{n}E\xi_k=\frac{\mu_n}{n}-p$$

故由车贝晓夫大数定律立刻推出贝努里大数定律.显然,贝努里大数定律也可以通过车贝晓夫不等式直接加以证明:

$$P\left\{\left|\frac{\mu_n}{n}-p\right|\geqslant\varepsilon\right\}\leqslant\frac{1}{\varepsilon^2}D\left(\frac{\mu_n}{n}\right)=\frac{1}{n\varepsilon^2}D\xi_i\leqslant\frac{1}{4n\varepsilon^2} \tag{18}$$

普阿松大数定律 如果在一个独立试验序列中,事件 A 在第 k 次试验中出现的概率等于 p_k,以 μ_n 记在前 n 次试验中事件 A 出现的次数,则对任意 $\varepsilon>0$,都有

$$\lim_{n\to\infty}P\left\{\left|\frac{\mu_n}{n}-\frac{p_1+p_2+\cdots+p_n}{n}\right|<\varepsilon\right\}=1 \tag{19}$$

[证明] 定义 ξ_k 为第 k 次试验中事件 A 出现的次数,则

$$E\xi_k=p_k, \qquad D\xi_k=p_k(1-p_k)\leqslant\frac{1}{4}$$

再用车贝晓夫大数定律立刻可以推出(19).

马尔可夫(Марков)注意到在车贝晓夫的论证中,只要

$$\frac{1}{n^2}D\left(\sum_{k=1}^{n}\xi_k\right)\longrightarrow 0 \tag{20}$$

则大数定律就能成立，通常称条件(20)为马尔可夫条件.

马尔可夫大数定律 对于随机变量序列$\xi_1, \xi_2, \cdots, \xi_n, \cdots$，若(20)成立，则对任意$\varepsilon>0$，均有(16).

车贝晓夫大数定律显然可由马尔可夫大数定律推出；更重要的是马尔可夫大数定律已经没有任何关于独立性的假定. 研究相依随机变量序列的大数定律是近代概率论的课题之一，但是这已超出我们讨论的范围.

3. 大数定律的重要意义

人们积累的大量经验告诉我们，具有很接近于1的概率的随机事件在一次试验中是几乎一定要发生；同样，概率很小的事件在一次试验中可以看作是实际不可能事件. 至于概率小到何种程度才能看作是实际不可能事件则要视事件的重要性而定，有百分之一可能性含有染菌的药物是应该废弃的，但含有百分之一次品的钮扣则问题还不太大.

因此在实际工作及一般理论问题中，概率接近于1或0的事件具有重大意义，概率论的基本问题之一就是要建立概率接近于1或0的规律；特别是大量独立或弱相关因素累积结果所发生的规律. 大数定律就是这种概率论命题中最重要的一个.

观察个别现象时是连同一切个别的特性来观察的. 这些个别的特性往往蒙蔽了事物的规律性. 在大量观察中个别因素的影响将相互抵消而使总体稳定. 例如，虽然每个气体分子的运动带有很大的随机性，但是作为气体平均特征的压力、温度等却是稳定的，大数定律说明了这种稳定性.

贝努里大数定律建立了在大量重复独立试验中事件出现频率的稳定性，正因为这种稳定性，概率的概念才有客观意义. 贝努里

大数定律还提供了通过试验来确定事件概率的方法，既然频率$\frac{\mu_n}{n}$与概率p有较大偏差的可能性很小，那么我们便可以通过做试验确定某事件发生的频率并把它作为相应概率的估计，这种方法称为**参数估计**，它是数理统计中的主要研究课题之一，参数估计的重要理论基础之一就是大数定律.

三、德莫哇佛-拉普拉斯极限定理

大数定律只断言$P\left\{\left|\frac{\mu_n}{n}-p\right|\geqslant\varepsilon\right\}$当$n\to\infty$时趋于0，也即$\frac{\mu_n}{n}$接近于$p$，而德莫哇佛-拉普拉斯极限定理则给出$\mu_n$的渐近分布的更精确表述. 下面定理给出了两个结果. 第一个结果提供了$P\{\mu_n=k\}$的渐近表达式，这类结果一般称为**局部极限定理**. 第二个结果给出了标准化随机变量$\frac{\mu_n-np}{\sqrt{npq}}$的渐近分布，称为**积分极限定理**，它是一般中心极限定理的特例. 这两个结果既有区别也有联系.

定理(德莫哇佛-拉普拉斯) 若μ_n是n次贝努里试验中事件A出现的次数，$0<p<1$，则对任意有限区间$[a,b]$：

(i) 当$a\leqslant x_k\equiv\frac{k-np}{\sqrt{npq}}\leqslant b$及$n\to\infty$时，一致地有

$$P\{\mu_n=k\}\div\left(\frac{1}{\sqrt{npq}}\cdot\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x_k^2}\right)\longrightarrow 1 \tag{21}$$

(ii) 当$n\to\infty$时，一致地有

$$P\left\{a\leqslant\frac{\mu_n-np}{\sqrt{npq}}<b\right\}\longrightarrow\int_a^b\varphi(x)dx \tag{22}$$

其中$\varphi(x)=\frac{1}{\sqrt{2\pi}}e^{-x^2/2}$，$(-\infty<x<\infty)$.

［证明］ 先证局部极限定理，我们将给出一个比(21)更为精确的渐近式.

因 x_k 只能在有限区间$[a, b]$中取值，故当 $n\to\infty$时，

$$k=np+x_k\sqrt{npq}\longrightarrow\infty \tag{23}$$

$$j\equiv n-k=nq-x_k\sqrt{npq}\longrightarrow\infty \tag{24}$$

由斯特灵(Stirling)公式:

$$m!=\sqrt{2\pi m}m^m e^{-m}e^{\theta_m}\qquad\left(0<\theta_m<\frac{1}{12m}\right) \tag{25}$$

可知

$$\begin{aligned}P\{\mu_n=k\}&=\frac{n!}{k!j!}p^kq^j\\&=\frac{\sqrt{2\pi n}n^ne^{-n}}{\sqrt{2\pi k}k^ke^{-k}\sqrt{2\pi j}j^je^{-j}}p^kq^je^{\theta_n-\theta_k-\theta_j}\\&=\frac{1}{\sqrt{2\pi}}\sqrt{\frac{n}{kj}}\left(\frac{np}{k}\right)^k\left(\frac{nq}{j}\right)^je^{\theta}\end{aligned} \tag{26}$$

其中 $\theta=\theta_n-\theta_k-\theta_j$，因此

$$|\theta|<\frac{1}{12}\left(\frac{1}{n}+\frac{1}{k}+\frac{1}{j}\right) \tag{27}$$

下面我们来估计(26)，由(23)及(24)

$$\begin{aligned}\sqrt{\frac{n}{kj}}&=\left(\frac{kj}{n}\right)^{-\frac{1}{2}}=\left[\frac{(np+x_k\sqrt{npq})(nq-x_k\sqrt{npq})}{n}\right]^{-\frac{1}{2}}\\&=\left[\frac{np\left(1+x_k\sqrt{\frac{q}{np}}\right)\cdot nq\left(1-x_k\sqrt{\frac{p}{nq}}\right)}{n}\right]^{-\frac{1}{2}}\\&=\frac{1}{\sqrt{npq}}\left[\left(1+x_k\sqrt{\frac{q}{np}}\right)\left(1-x_k\sqrt{\frac{p}{nq}}\right)\right]^{-\frac{1}{2}}\\&=\frac{1}{\sqrt{npq}}\left[1+\frac{x_k(q-p)}{\sqrt{npq}}+O\left(\frac{1}{n}\right)\right]^{-\frac{1}{2}}\\&=\frac{1}{\sqrt{npq}}\left[1-\frac{x_k(q-p)}{2\sqrt{npq}}+O\left(\frac{1}{n}\right)\right]\end{aligned} \tag{28}$$

这里我们用 $O\left(\frac{1}{n}\right)$表示与$\frac{1}{n}$同阶的无穷小量。

由(23)及(24)知

$$\frac{k}{np}=1+x_k\sqrt{\frac{q}{np}},\quad \frac{j}{nq}=1-x_k\sqrt{\frac{p}{nq}} \tag{29}$$

我们将利用展开式

$$\ln(1+x)=x-\frac{x^2}{2}+\frac{x^3}{3}+\cdots \tag{30}$$

来对它们进行估计．这个展开式当$-1<x\leqslant 1$时收敛，但只有对绝对值很小的那种x值才收敛得快。现在，当n充分大时，$x_k\sqrt{\frac{q}{np}}$及$x_k\sqrt{\frac{p}{nq}}$都很小(因此，当$p=0$及$p=1$时不能用；此外，当p或q很小时，渐近展开式引起的误差也较大，这时我们已推荐用普阿松逼近公式)。所以

$$\begin{aligned}\ln\left(\frac{np}{k}\right)^k\left(\frac{nq}{j}\right)^j &= -k\ln\frac{k}{np}-j\ln\frac{j}{nq}\\ &= -(np+x_k\sqrt{npq})\left[x_k\sqrt{\frac{q}{np}}-\frac{1}{2}\frac{qx_k^2}{np}\right.\\ &\quad\left.+\frac{1}{3}\left(\frac{q}{np}\right)^{\frac{3}{2}}x_k^3+O\left(\frac{1}{n^2}\right)\right]\\ &\quad+(nq-x_k\sqrt{npq})\left[x_k\sqrt{\frac{p}{nq}}+\frac{1}{2}\frac{px_k^2}{nq}\right.\\ &\quad\left.+\frac{1}{3}\left(\frac{p}{nq}\right)^{\frac{3}{2}}x_k^3+O\left(\frac{1}{n^2}\right)\right]\end{aligned}$$

对上式化简并略去高阶无穷小，可以得到

$$\ln\left(\frac{np}{k}\right)^k\left(\frac{nq}{j}\right)^j=-\frac{1}{2}x_k^2+\frac{x_k^3(q-p)}{6\sqrt{npq}}+O\left(\frac{1}{n}\right) \tag{31}$$

因此

$$\left(\frac{np}{k}\right)^k\left(\frac{nq}{j}\right)^j=e^{-\frac{1}{2}x_k^2}\cdot e^{x_k^3(q-p)/6\sqrt{npq}}\cdot e^{O(1/n)}$$

$$=e^{-\frac{1}{2}x_k^2}\left[1+\frac{x_k^3(q-p)}{6\sqrt{npq}}+O\left(\frac{1}{n}\right)\right]$$

$$\cdot\left[1+O\left(\frac{1}{n}\right)\right]$$

$$=e^{-\frac{1}{2}x_k^2}\left[1+\frac{x_k^3(q-p)}{6\sqrt{npq}}+O\left(\frac{1}{n}\right)\right] \tag{32}$$

把(28)及(32)代入(26)并注意到(27)可知:

$$P\{\mu_n=k\}=\frac{1}{\sqrt{2\pi}}\cdot\frac{1}{\sqrt{npq}}e^{-\frac{1}{2}x_k^2}\left[1+\frac{x_k^3(q-p)}{6\sqrt{npq}}\right.$$

$$\left.-\frac{x_k(q-p)}{2\sqrt{npq}}+O\left(\frac{1}{n}\right)\right]$$

$$=\frac{1}{\sqrt{npq}}\cdot\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x_k^2}\left[1+\frac{(x_k^3-3x_k)(q-p)}{6\sqrt{npq}}\right.$$

$$\left.+O\left(\frac{1}{n}\right)\right] \tag{33}$$

取其第一项即得(21)，因此我们已证得了局部极限定理．下面证明积分极限定理部分．

$$P\left\{a\leqslant\frac{\mu_n-np}{\sqrt{npq}}<b\right\}$$

$$=P\{np+a\sqrt{npq}\leqslant\mu_n<np+b\sqrt{npq}\}$$

$$=\sum_{k=k_1}^{k_2}P\{\mu_n=k\} \tag{34}$$

上式中k_1为不小于$np+a\sqrt{npq}$的最小整数，k_2为小于$np+b\sqrt{npq}$的最大整数，由局部极限定理知当n充分大时，对任给$\varepsilon>0$，有

$$P\{\mu_n=k\}=\frac{1}{\sqrt{npq}}(\varphi(x_k)+e_k),\quad |e_k|<\varepsilon$$

$$(k=k_1, k_1+1, \cdots, k_2)$$

代入(34)式得到

$$P\left\{a \leqslant \frac{\mu_n - np}{\sqrt{npq}} < b\right\}$$

$$= \sum_{k=k_1}^{k_2} \frac{1}{\sqrt{npq}} \varphi(x_k) + \sum_{k=k_1}^{k_2} \frac{e_k}{\sqrt{npq}}$$

因为有

$$\left| \sum_{k=k_1}^{k_2} \frac{e_k}{\sqrt{npq}} \right| \leqslant \frac{1}{\sqrt{npq}} (k_2 - k_1 + 1) e \leqslant \frac{(b-a)\sqrt{npq}+1}{\sqrt{npq}} e$$

故当 $n \to \infty$时，注意 x_k 的增量为$\frac{1}{\sqrt{npq}}$，就得到

$$P\left\{a \leqslant \frac{\mu_n - np}{\sqrt{npq}} < b\right\} \longrightarrow \int_a^b \varphi(x) dx$$

定理到此完全证毕.

利用$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/2} dx = 1$，不难证明在积分极限定理中，当 $a = -\infty$, $b = +\infty$时仍然成立.

四、德莫哇佛-拉普拉斯极限定理的一些应用

德莫哇佛-拉普拉斯极限定理虽然是作为二项分布的近似而推导出来的，但是它的重要性远远超出数值计算的范围. 对它的各种推广形式的深入讨论将在以后几节进行，这里我们先介绍一些它的具体应用，其中有些解决了第一、二章遗留的问题.

[导出贝努里大数定律] 积分极限定理断言，μ_n的分布渐近于正态分布$N(np, npq)$，因此贝努里试验中事件A出现的频率 $\frac{\mu_n}{n}$ 的分布渐近于正态分布 $N\left(p, \frac{pq}{n}\right)$. 这里可以想象，当 $n \to \infty$ 时，$\frac{\mu_n}{n}$ 的分布会收敛于退化分布

$$I(x-p)=\begin{cases}0, & x\leqslant p\\ 1, & x>p\end{cases}$$

这正是贝努里大数定律所确定的事实．下面来严格证明．

因为 ε 固定，对任意正数 l，只要 n 足够大，就有

$$l\sqrt{npq}<\varepsilon n$$

因此

$$\left\{\left|\frac{\mu_n-np}{\sqrt{npq}}\right|<l\right\}\subset\left\{\left|\frac{\mu_n-np}{n}\right|<\varepsilon\right\}$$

所以对大的 n，

$$P\left\{\left|\frac{\mu_n-np}{\sqrt{npq}}\right|<l\right\}\leqslant P\left\{\left|\frac{\mu_n-np}{n}\right|<\varepsilon\right\}$$

由积分极限定理，当 $n\to\infty$ 时，上式左边收敛于

$$\frac{1}{\sqrt{2\pi}}\int_{-l}^{l}e^{-x^2/2}dx$$

对于任给 $\delta>0$，可以选 l，使得上面积分值大于 $1-\delta$，因此对充分大的 n 有

$$P\left\{\left|\frac{\mu_n}{n}-p\right|<\varepsilon\right\}>1-\delta$$

这就证明了贝努里大数定律，从这里也可看出，积分极限定理比大数定律更精确．

［用频率估计概率时误差的估计］ 由积分极限定理

$$P\left\{\left|\frac{\mu_n}{n}-p\right|<\varepsilon\right\}=P\left\{-\varepsilon\sqrt{\frac{n}{pq}}<\frac{\mu_n-np}{\sqrt{npq}}<\varepsilon\sqrt{\frac{n}{pq}}\right\}$$

$$\approx\Phi\left(\varepsilon\sqrt{\frac{n}{pq}}\right)-\Phi\left(-\varepsilon\sqrt{\frac{n}{pq}}\right)=2\Phi\left(\varepsilon\sqrt{\frac{n}{pq}}\right)-1 \qquad (35)$$

这个关系式可用来解决许多计算问题．

第一类问题是已知 n, p, ε，求概率 $P\left\{\left|\frac{\mu_n}{n}-p\right|<\varepsilon\right\}$；这时只要

利用(35)并查正态分布函数 $\Phi(x)$ 的数值表就可解决，这类问题在二项分布计算中经常会遇到。

第二类问题是要使$\frac{\mu_n}{n}$与 p 的差异不大于定数 ε 的概率不小于预先给的数 β，问最少应做多少次试验？这时只需要求满足下式的最小 n，

$$2\Phi\left(\varepsilon\sqrt{\frac{n}{pq}}\right)-1\geqslant\beta$$

这也可通过查表求得.

第三类问题已知 n 及 β，求 ε，这类问题是在进行误差估计时提出来的. 解法如下：先找 x_β 使

$$2\Phi(x_\beta)-1=\beta$$

这时

$$\varepsilon=x_\beta\sqrt{\frac{pq}{n}}$$

即为所求. 若 p 不知道，则利用 $pq\leqslant\frac{1}{4}$，有下列估计式

$$\varepsilon\leqslant\frac{x_\beta}{2\sqrt{n}}$$

这类估计在蒙特卡洛方法中很有用处.

[局部极限定理在二项分布计算中的应用]局部极限定理给出

$$\binom{n}{k}p^kq^{n-k}\div\left(\frac{1}{\sqrt{npq}}\cdot\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{k-np}{\sqrt{npq}}\right)^2}\right)\longrightarrow1 \tag{36}$$

这个事实由图1表示出来，其中阶梯函数给出概率$\binom{n}{k}p^kq^{n-k}$，而粗线则给出对应的正态分布密度函数曲线.

我们来利用局部极限定理对$\binom{n}{k}p^kq^{n-k}$进行近似计算.

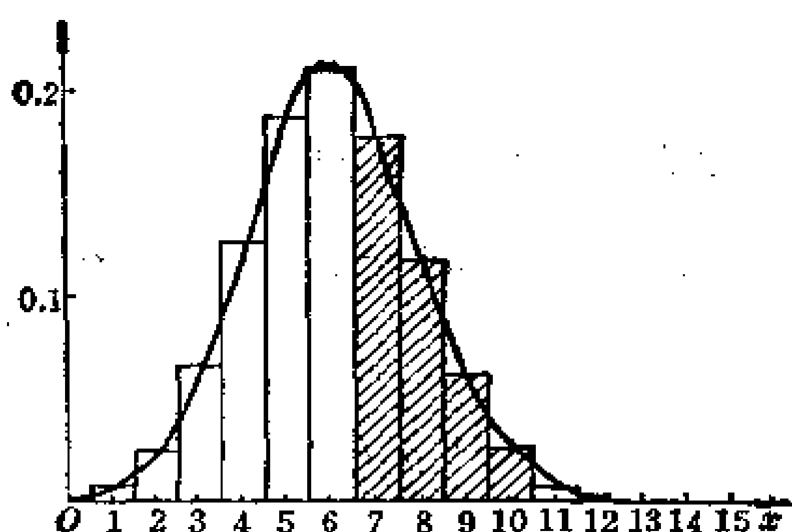


图1　二项分布的正态逼近

因为
$$\varphi(x)=\frac{1}{\sqrt{2\pi}}e^{-x^2/2},\ -\infty<x<\infty \tag{37}$$

有专门数值表，当 n 较大时，对二项分布的计算可用下列近似式：

$$\binom{n}{k}p^k q^{n-k}\approx\frac{1}{\sqrt{npq}}\varphi(x_k) \tag{38}$$

其中 $x_k=\frac{k-np}{\sqrt{npq}}$. 用这种方法计算当然会有误差，但实践证明（也能作理论分析），当 p 不太接近于 0 或 1，而 n 又不太小时，都能得到良好的结果.

下面就用这种办法来解决第二章 § 4 中提出的一些二项分布计算问题.

例 3 中要求计算 $b(5;\ 500,\ 0.01)=\binom{500}{5}(0.01)^5\cdot(0.99)^{495}$，利用(38)

$$\begin{aligned}&b(5;500,0.01)\\&\approx\frac{1}{\sqrt{500\times0.01\times0.99}}\varphi\left(\frac{5-500\times0.01}{\sqrt{500\times0.01\times0.99}}\right)\\&=\frac{1}{\sqrt{4.95}}\varphi(0)=\frac{1}{\sqrt{4.95}}\times0.398942=0.17931\end{aligned}$$

第二章中已指出 $b(5;\ 500, 0.01)$ 的精确值是 0.17635. 这里 $p=0.01$ 很接近于 0, 我们用近似计算得到的数值的有效性还达到小数点后两位, 由此可见用局部极限定理计算的近似程度是很好的.

再来计算例 4(1), 那里要求计算 $b(40;10000,\ 0.005)$, 利用 (38)

$$b(40;10000,0.005)$$

$$\approx\frac{1}{\sqrt{10000\times0.005\times0.995}}\varphi\left(\frac{40-10000\times0.005}{\sqrt{10000\times0.005\times0.995}}\right)$$

$$=\frac{1}{7.05}\varphi(1.4)=\frac{1}{7.05}\times0.1497=0.0212$$

而精确值为 0.0214, 准确度甚高.

[积分极限定理在二项分布计算中的应用] 积分极限定理给出

$$P\left\{a\leqslant\frac{\mu_n-np}{\sqrt{npq}}<b\right\}\longrightarrow\Phi(b)-\Phi(a) \tag{39}$$

其中 $\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-t^2/2}dt$ 是正态分布函数, 利用它可以计算形如 $P\{k_1\leqslant\mu_n\leqslant k_2\}$ 的概率.

因为当 n 充分大时, $b(k;n,p)$ 即使对于最可能成功次数, 其数值也很小, 这时讨论成功次数等于某数 k, 即求 $P\{\mu_n=k\}$ 便没有太大意思; 相反, 计算 μ_n 的值落在某一区间 $[k_1, k_2]$ 的概率倒十分重要, 因此积分极限定理在二项分布近似计算中更为有用.

我们有

$$P\{k_1\leqslant\mu_n\leqslant k_2\}=P\left\{\frac{k_1-np}{\sqrt{npq}}\leqslant\frac{\mu_n-np}{\sqrt{npq}}\leqslant\frac{k_2-np}{\sqrt{npq}}\right\}$$

$$\approx\Phi\left(\frac{k_2-np}{\sqrt{npq}}\right)-\Phi\left(\frac{k_1-np}{\sqrt{npq}}\right) \tag{40}$$

当 p 不太接近 0 或 1，而 n 又不太小时，用这个近似式能得到良好的结果.

例如，我们来完成第二章 § 4 中的一些计算. 例 4(2) 要计算 $P\{\mu\leqslant 70\}$，利用(40)

$$
\begin{aligned}
&P\{\mu\leqslant 70\}\\
&\approx\Phi\left(\frac{70-10000\times0.005}{\sqrt{10000\times0.005\times0.995}}\right)-\Phi\left(\frac{-10000\times0.005}{\sqrt{10000\times0.005\times0.995}}\right)\\
&=\Phi(2.84)-\Phi(-7.09)=\Phi(2.84)-[1-\Phi(7.09)]=0.997
\end{aligned}
$$

这里 $\Phi(7.09)$ 取为 1，此外计算中还用到 $\Phi(-x)=1-\Phi(x)$，这由 $\varphi(x)$ 的对称性即得.

再来计算例 5，其中 $p=\frac{1}{2}$，$n=5.4\times10^{22}$.

$$
P\{|\mu-np|>2.7\times10^{12}\}=P\left\{\frac{|\mu-np|}{\sqrt{npq}}>\frac{2.7\times10^{12}}{\sqrt{5.4\times10^{22}\times\frac{1}{4}}}\right\}
$$

$$
=P\left\{\frac{|\mu-np|}{\sqrt{npq}}>2.32\times10\right\}\approx\frac{2}{\sqrt{2\pi}}\int_{23.2}^{\infty}e^{-t^2/2}\,dt
$$

这个数值非常小，从 $\Phi(x)$ 的数值表中不能找到，但我们可以用下面方法对它进行估计.

由于对 $z>0$ 有

$$
\int_z^{\infty}e^{-t^2/2}dt<\frac{1}{z}\int_z^{\infty}t\,e^{-t^2/2}dt=\frac{1}{z}e^{-z^2/2} \tag{41}
$$

利用它可以得到

$$
\frac{2}{\sqrt{2\pi}}\int_{23.2}^{\infty}e^{-t^2/2}dt<\frac{2}{\sqrt{2\pi}}\cdot\frac{1}{23.2}e^{-2.7\times100}<10^{-100}
$$

这概率非常小.

现在我们能够解决第一章 § 1 中提出的用电问题了. 该问题可以重新叙述如下：

某车间有 200 台车床，它们独立地工作着，开工率各为 0.6，开

工时耗电各为1千瓦，问供电所至少要供给这个车间多少电力才能以99.9%的概率保证这个车间不会因供电不足而影响生产.

这里的叙述已把问题数学模型化，提出了各车床工作的独立性假定，并把所谓能正常生产的要求加以明确化，这是在用数学解决生产实际问题时必须经过的一步.

现在，这成为试验次数 $n=200$ 的贝努里试验，若把某台车床在工作看作成功，则出现成功的概率为0.6. 记某时在工作着的车床数为 ξ，则 ξ 是随机变量，服从 $p=0.6$ 的二项分布. 问题是要求 r，使

$$P\{\xi\leqslant r\}=\sum_{k=0}^{r}\binom{200}{k}(0.6)^k(0.4)^{200-k}\geqslant 0.999 \tag{42}$$

我们可以利用积分极限定理计算这个概率.

$$\begin{aligned}&\sum_{k=0}^{r}\binom{200}{k}(0.6)^k(0.4)^{200-k}\\&\approx\Phi\left(\frac{r-200\times0.6}{\sqrt{200\times0.6\times0.4}}\right)-\Phi\left(\frac{-200\times0.6}{\sqrt{200\times0.6\times0.4}}\right)\\&=\Phi\left(\frac{r-120}{\sqrt{48}}\right)-\Phi(-17.32)\doteq\Phi\left(\frac{r-120}{\sqrt{48}}\right)\geqslant0.999\end{aligned}$$

查表得

$$\frac{r-120}{\sqrt{48}}=3.1$$

所以

$$r=141$$

这个结果表明 $P\{\xi\leqslant141\}\geqslant0.999$，所以我们若供电141千瓦，那么由于供电不足而影响生产的可能性小于0.001，相当于在8小时工作中有半分钟受影响，这在一般工厂中是允许的. 当然不同的生产单位，可能提出不同的要求，那么我们可以改变(42)右端的概

率值，但是方法还是同样的.

现在也能对高尔顿板这个试验作理论解释了．读者可能早已想到，近似小球高度的曲线是正态分布密度函数曲线，这完全正确．事实上，高尔顿板可以看作是贝努里试验的一个实验模型．如果我们把小球碰到钉子看作是一次试验，而把从右边落下算是成功，当然从左边落下就算是失败，这时就有了一次 $p=\frac{1}{2}$ 的贝努里试验．小球从顶端到底层共需要经过 n 排钉子，这就相当于一个 n 次贝努里试验，剩下的只是要说明为什么高度曲线会是正态分布密度函数曲线，这个问题留给读者思考．

§2. 收 敛 性

从下节开始，我们将把在贝努里试验场合建立的极限定理推广到更为一般的场合，本节为此准备了必要的概念与工具．

从上节的讨论中我们已经看到，概率论的极限定理研究的是随机变量序列与分布函数序列的某种收敛性，下面我们将给这些收敛性以明确定义并讨论它们的有关性质．这些结果对于深入研究概率论也有着独立的重要性．

特征函数是研究极限定理的有力工具，从上章的讨论中我们已经知道，它与分布函数互相唯一确定，本节中我们将证明这种对应还具有某种连续性，这些性质决定了特征函数在极限定理研究中的特殊地位．顺便还得到特征函数的充要条件，这个结果在平稳随机过程的研究中有基本的重要性．本节定理的证明都较长，初学时可以略去，后面用到的是结论．

一、分布函数弱收敛

中心极限定理讨论的是分布函数列收敛于正态分布．事实上，在德莫哇佛-拉普拉斯积分极限定理中，若记

$$F_n(x)=P\left\{\frac{\mu_n-np}{\sqrt{npq}}<x\right\}$$

则定理的结论可以表述为

$$F_n(x)\to\Phi(x)$$

这正是一个分布函数列$\{F_n(x)\}$收敛于某一个分布函数$\Phi(x)$，这种收敛对每一点x都成立。

这个考察对于我们引进一般分布函数列的收敛性定义很有帮助，在给出定义前，我们再来看一个例子。

[例1] 令

$$F_n(x)=\begin{cases}0, & x\leqslant-\frac{1}{n}\\ 1, & x>-\frac{1}{n}\end{cases} \tag{1}$$

这是一个退化分布，它可以解释为一个单位质量全部集中在$x=-\frac{1}{n}$这一点的分布．当$n\to\infty$时，我们自然认为$\{F_n(x)\}$应该收敛于一个单位质量全部集中在$x=0$这一点的分布，即

$$F(x)=\begin{cases}0, & x\leqslant0\\ 1, & x>0\end{cases}$$

但是，$F_n(0)\equiv1$，而$F(0)=0$，显然，$F_n(0)\not\to F(0)$．因此看来要求分布函数列在所有的点都收敛到极限分布函数是太严了．上例中不收敛的点是极限分布函数$F(x)$的不连续点．

定义1 对于分布函数列$\{F_n(x)\}$，如果存在一个函数$F(x)$使

$$\lim_{n\to\infty}F_n(x)=F(x) \tag{2}$$

在$F(x)$的每一连续点上都成立，则称$F_n(x)$**弱收敛于**$F(x)$，并记为$F_n(x)\xrightarrow{W}F(x)$．

这样得到的极限函数是一个有界的非降的函数，我们也可以

选得它是左连续的，但是下例说明，它不一定是一个分布函数.

［例 2］ 取

$$F_n(x)=\begin{cases}0, & x\leqslant n\\ 1, & x>n\end{cases} \tag{3}$$

显然$\lim\limits_{n\to\infty}F_n(x)=0$对一切$x$成立，但$F(x)\equiv 0$不是分布函数.

当然，若已知分布函数列$\{F_n(x)\}$弱收敛于分布函数$F(x)$及$G(x)$，则$F(x)=G(x)$对一切x成立.

我们希望能得到一个分布函数列弱收敛到一个分布函数的充要条件，为此先建立一些重要的分析结果，这些结果对一般的有界非降函数列都成立，它们的弱收敛概念类似地定义.

引理 1 设$\{F_n(x)\}$是实变量x的非降函数列，D是R^1上的稠密集. 若对于D中的所有点，序列$\{F_n(x)\}$收敛于$F(x)$，则对$F(x)$的一切连续点x有

$$\lim_{n\to\infty}F_n(x)=F(x) \tag{4}$$

［证明］ 设x是任意点，选$x'\in D$，$x''\in D$，使$x'\leqslant x\leqslant x''$，由非降性知

$$F_n(x')\leqslant F_n(x)\leqslant F_n(x'')$$

因此

$$F(x')\leqslant\varliminf_{n\to\infty}F_n(x)\leqslant\varlimsup_{n\to\infty}F_n(x)\leqslant F(x'')$$

因为D在R^1上稠密，故

$$F(x-0)\leqslant\varliminf_{n\to\infty}F_n(x)\leqslant\varlimsup_{n\to\infty}F_n(x)\leqslant F(x+0)$$

所以对于$F(x)$的连续点x，成立(4).

下面证明海莱(Helly)的两个重要定理.

定理 1（海莱第一定理） 任一一致有界的非降函数列$\{F_n(x)\}$中必有一子序列$\{F_{n_k}(x)\}$弱收敛于某一有界的非降函数

$F(x)$.

[证明] 任取 R^1 上的一个到处稠密的可数点集 D, 下面我们就取有理数全体, 并排列为 $r_1, r_2, \cdots, r_m, \cdots$. 对于序列$\{F_n(r_1)\}$, 这是一个有界的实数序列, 故必包含一收敛于某极限 $G(r_1)$的子序列$\{F_{1,n}(r_1)\}$, 即

$$\lim_{n\to\infty} F_{1,n}(r_1) = G(r_1)$$

现在考虑序列$\{F_{1,n}(r_2)\}$, 同样由于有界性, 在其中存在子序列$\{F_{2,n}(r_2)\}$收敛于某一值 $G(r_2)$. 这时, 同时成立着

$$\lim_{n\to\infty} F_{2,n}(r_1) = G(r_1), \quad \lim_{n\to\infty} F_{2,n}(r_2) = G(r_2)$$

继续这样做, 可得序列$\{F_{m,n}(x)\}$, 使

$$\lim_{n\to\infty} F_{m,n}(r_k) = G(r_k), \quad k = 1, 2, \cdots, m \tag{5}$$

同时成立.

这样, 我们得到了$\{F_n(x)\}$的如下子序列

$$\begin{array}{l} F_{1,1}(x), F_{1,2}(x), F_{1,3}(x), \cdots, F_{1,n}(x), \cdots \\ F_{2,1}(x), F_{2,2}(x), F_{2,3}(x), \cdots, F_{2,n}(x), \cdots \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ F_{m,1}(x), F_{m,2}(x), F_{m,3}(x), \cdots, F_{m,n}(x), \cdots \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \end{array} \tag{6}$$

这里每行都是前一行的子序列, 而且它们具有性质(5). 选取这个阵列的对角线元素 $F_{n,n}(x)$构成新序列$\{F_{n,n}(x)\}$, 由于它是从$\{F_{1,n}(x)\}$分出来的, 故$\lim\limits_{n\to\infty} F_{n,n}(r_1) = G(r_1)$. 其次, 除第一项外, 它是由$\{F_{2,n}(x)\}$分出来的, 故$\lim\limits_{n\to\infty} F_{n,n}(r_2) = G(r_2)$. 一般地, 对任何固定的 k, 皆有 $\lim\limits_{n\to\infty} F_{n,n}(r_k) = G(r_k)$, 因此对一切有理数 r,

$$\lim_{n\to\infty} F_{n,n}(r) = G(r) \tag{7}$$

这里的 $G(r)$是定义在有理数上的函数, 它也是有界与非降的.

对一切 $x\in R^1$，定义

$$F(x)=\sup_{r_k\leqslant x}G(r_k)$$

这函数在一切有理数上与 $G(x)$ 相等，它显然也是有界与非降的。由引理1知

$$\lim_{n\to\infty}F_{n,n}(x)=F(x) \tag{8}$$

对 $F(x)$ 的一切连续点成立，这就证明了定理1。通常形象地称这个定理的证明方法为**对角线法**。

极限函数 $F(x)$ 不一定左连续，但总可以改变它不连续点上的值使之左连续，这样的改变显然不影响(8)的成立。

定理2（海莱第二定理） 设 $f(x)$ 是 $[a,b]$ 上的连续函数，又 $\{F_n(x)\}$ 是在 $[a,b]$ 上弱收敛于函数 $F(x)$ 的一致有界非降函数序列，且 a 和 b 是 $F(x)$ 的连续点，则

$$\lim_{n\to\infty}\int_a^b f(x)dF_n(x)=\int_a^b f(x)dF(x)$$

［证明］ 由函数 $f(x)$ 的连续性推知，对任意正数 ε，总可以找到一种分割，把区间 $[a,b]$ 分为 $[x_0,x_1],[x_1,x_2],\cdots,[x_{N-1},x_N]$（其中 $x_0=a,x_N=b$）等 N 个小区间，使得当 $x\in[x_k,x_{k+1}]$ 时，$|f(x)-f(x_k)|<\varepsilon$。利用这种情况，我们能导入一个辅助函数 $f_\varepsilon(x)$，它只取有限个值，并且当 $x_k<x<x_{k+1}$ 时，$f_\varepsilon(x)=f(x_k)$。

这样显然对 $a\leqslant x\leqslant b$ 的一切 x 皆有不等式

$$|f(x)-f_\varepsilon(x)|<\varepsilon \tag{9}$$

在此我们可预先选取分点 $x_1,x_2,\cdots,x_{N-1}$，使它们是 $F(x)$ 的连续点。因为 $\{F_n(x)\}$ 弱收敛于 $F(x)$，故当 n 充分大时，在此 $N-1$ 个分点及 x_0,x_N 上成立不等式

$$|F(x_k)-F_n(x_k)|<\frac{\varepsilon}{MN} \tag{10}$$

这里 M 是 $|f(x)|$ 在区间 $a\leqslant x\leqslant b$ 中的最大值。显然，

$$
\begin{aligned}
&\left|\int_a^b f(x)dF(x)-\int_a^b f(x)dF_n(x)\right| \\
&\leqslant\left|\int_a^b f(x)dF(x)-\int_a^b f_\varepsilon(x)dF(x)\right| \\
&\quad+\left|\int_a^b f_\varepsilon(x)dF(x)-\int_a^b f_\varepsilon(x)dF_n(x)\right| \\
&\quad+\left|\int_a^b f_\varepsilon(x)dF_n(x)-\int_a^b f(x)dF_n(x)\right| \qquad (11)
\end{aligned}
$$

由于(9),

$$
\left|\int_a^b f(x)dF(x)-\int_a^b f_\varepsilon(x)dF(x)\right|\leqslant\varepsilon[F(b)-F(a)] \qquad (12)
$$

$$
\begin{aligned}
&\left|\int_a^b f_\varepsilon(x)dF_n(x)-\int_a^b f(x)dF_n(x)\right| \\
&\leqslant\varepsilon[F_n(b)-F_n(a)] \qquad (13)
\end{aligned}
$$

而由(9), (10)可知

$$
\begin{aligned}
&\left|\int_a^b f_\varepsilon(x)dF(x)-\int_a^b f_\varepsilon(x)dF_n(x)\right| \\
&=\left|\sum_{k=0}^{N-1}f(x_k)[F(x_{k+1})-F(x_k)]-\sum_{k=0}^{N-1}f(x_k)[F_n(x_{k+1})-F_n(x_k)]\right| \\
&=\left|\sum_{k=0}^{N-1}f(x_k)[F(x_{k+1})-F_n(x_{k+1})]-\sum_{k=0}^{N-1}f(x_k)[F(x_k)-F_n(x_k)]\right| \\
&\leqslant N\left[M\frac{\varepsilon}{MN}+M\frac{\varepsilon}{MN}\right]=2\varepsilon \qquad (14)
\end{aligned}
$$

因此

$$
\begin{aligned}
&\left|\int_a^b f(x)dF(x)-\int_a^b f(x)dF_n(x)\right| \\
&\leqslant\varepsilon[F(b)-F(a)]+\varepsilon[F_n(b)-F_n(a)]+2\varepsilon
\end{aligned}
$$

由于$\{F_n(x)\}$的一致有界性，上式右边可以任意小，故定理得证。

定理3（拓广的海莱第二定理） 设$f(x)$在$(-\infty,\infty)$上有界

连续，又$\{F_n(x)\}$是$(-\infty, +\infty)$上弱收敛于函数$F(x)$的一致有界非降函数序列，且

$$\lim_{n\to\infty} F_n(-\infty)=F(-\infty), \lim_{n\to\infty} F_n(+\infty)=F(+\infty)$$

则

$$\lim_{n\to\infty}\int_{-\infty}^{\infty} f(x)dF_n(x)=\int_{-\infty}^{\infty} f(x)dF(x)$$

[证明] 设$A<0, B>0$，令

$$J_1=\left|\int_{-\infty}^{A} f(x)dF_n(x)-\int_{-\infty}^{A} f(x)dF(x)\right|$$

$$J_2=\left|\int_{A}^{B} f(x)dF_n(x)-\int_{A}^{B} f(x)dF(x)\right|$$

$$J_3=\left|\int_{B}^{\infty} f(x)dF_n(x)-\int_{B}^{\infty} f(x)dF(x)\right|$$

显然

$$\left|\int_{-\infty}^{\infty} f(x)dF_n(x)-\int_{-\infty}^{\infty} f(x)dF(x)\right|\leqslant J_1+J_2+J_3$$

由于$f(x)$是有界的，存在常数$M>0$，使$|f(x)|<M$. 又由于序列$\{F_n(x)\}$的一致有界性，只要A与B的绝对值充分大，并使A和B是$F(x)$的连续点，而n也取得充分大，则可使J_1, J_3小到预先给定的程度. 事实上

$$J_1\leqslant M[|F_n(A)-F_n(-\infty)|+|F(A)-F(-\infty)|]$$

$$J_3\leqslant M[|F_n(+\infty)-F_n(B)|+|F(+\infty)-F(B)|]$$

而按假定有

$$\lim_{n\to\infty} F_n(A)=F(A), \quad \lim_{n\to\infty} F_n(B)=F(B)$$

$$\lim_{n\to\infty} F_n(-\infty)=F(-\infty), \lim_{n\to\infty} F_n(+\infty)=F(+\infty)$$

故J_1, J_3可以任意小. 再根据定理2，只要n充分大，也可使J_2任意小，从而证得了定理.

二、连续性定理

下面我们将导出一个分布函数列弱收敛到一个极限分布的充要条件，这个结果同时说明了存在于分布函数与特征函数之间的一一对应是连续的，这个性质对于特征函数成为研究一些极限定理的主要工具有基本的重要性.

定理 4（正极限定理） 设分布函数列$\{F_n(x)\}$弱收敛于某一分布函数 $F(x)$，则相应的特征函数列$\{f_n(t)\}$收敛于特征函数 $f(t)$，且在 t 的任一有限区间内收敛是一致的.

［证明］ 函数 e^{itx} 在$-\infty<x<+\infty$上有界连续，而

$$f_n(t)=\int_{-\infty}^{\infty}e^{itx}dF_n(x)$$

$$f(t)=\int_{-\infty}^{\infty}e^{itx}dF(x)$$

因此由拓广的海莱第二定理即知当 $n\to\infty$时，有

$$f_n(t)\to f(t)$$

至于在 t 的每一有限区间内收敛的一致性，由拓广的海莱第二定理的证明就可看出.

定理 5（逆极限定理） 设特征函数列$\{f_n(t)\}$收敛于某一函数 $f(t)$，且 $f(t)$在 $t=0$ 连续，则相应的分布函数列$\{F_n(x)\}$弱收敛于某一分布函数 $F(x)$，而且 $f(t)$是 $F(x)$的特征函数.

［证明］ 由海莱第一定理，知必存在子序列$\{F_{n_k}(x)\}$弱收敛于某一非降函数 $F(x)$，且 $F(x)$可视为左连续的. 极限函数 $F(x)$显然满足 $F(-\infty)\geqslant 0$，$F(+\infty)\leqslant 1$，我们来证明 $F(x)$是分布函数. 否则，应有

$$\delta=F(+\infty)-F(-\infty)<1 \tag{15}$$

任取一正数 $\varepsilon<1-\delta$. 因 $f(t)$是特征函数列的极限，故 $f(0)=1$. 由于 $f(t)$在 $t=0$ 是连续的，故可选取充分小的正数 τ，使

$$\frac{1}{2\tau}\left|\int_{-\tau}^{\tau} f(t)dt\right|>1-\frac{\varepsilon}{2}>\delta+\frac{\varepsilon}{2}$$

η下定义.

)的分布函数

同时选取 $X\geqslant\frac{4}{\tau\varepsilon}$ 及 K，使 $k\geqslant K$ 时，

$$\delta_k=F_{n_k}(X)-F_{n_k}(-X)\leqslant\delta+\frac{\varepsilon}{4}$$

又因 $f_{n_k}(t)$ 是特征函数，那末

$$\int_{-\tau}^{\tau} f_{n_k}(t)dt=\int_{-\infty}^{\infty}\left[\int_{-\tau}^{\tau} e^{itx}dt\right]dF_{n_k}(x) \tag{17}$$

显然

$$\left|\int_{-\tau}^{\tau} e^{itx}dt\right|\leqslant 2\tau$$

还有，在 $|x|>X$ 时，

$$\left|\int_{-\tau}^{\tau} e^{itx}dt\right|=\left|\frac{2}{x}\sin\tau x\right|<\frac{2}{X} \tag{18}$$

因此

$$\left|\int_{-\tau}^{\tau} f_{n_k}(t)dt\right|\leqslant\left|\int_{|x|\leqslant X}\left(\int_{-\tau}^{\tau} e^{itx}dt\right)dF_{n_k}(x)\right|$$

$$+\left|\int_{|x|>X}\left(\int_{-\tau}^{\tau} e^{itx}dt\right)dF_{n_k}(x)\right|$$

$$<2\tau\delta_k+\frac{2}{X}$$

所以

$$\frac{1}{2\tau}\left|\int_{-\tau}^{\tau} f_{n_k}(t)dt\right|\leqslant\delta_k+\frac{1}{X\tau}\leqslant\delta+\frac{\varepsilon}{2}$$

令 $k\to\infty$，由控制收敛定理知

$$\frac{1}{2\tau}\left|\int_{-\tau}^{\tau} f(t)dt\right|\leqslant\delta+\frac{\varepsilon}{2}$$

这与(16)矛盾，因此(15)不成立，也即应有

$$F(-\infty)=0,\quad F(+\infty)=1$$

布函数，再由定理 4 推知 $f(t)$ 是 $F(x)$ 的特征函数。

……$\{F_n(x)\}$ 也弱收敛于同一分布函数 $F(x)$。如其不然，一定存在 $F(x)$ 的一个连续点 x_0 使 $\{F_n(x_0)\}$ 不收敛于 $F(x_0)$。这时可从 $\{F_n(x_0)\}$ 中选取一个收敛的子序列 $\{F_{m_k}(x_0)\}$，其极限 $F^*(x_0)\neq F(x_0)$。根据海莱第一定理，一定可以选取 $\{F_{m_k}(x)\}$ 的一个子序列 $\{F_{m_{k_l}}(x)\}$ 弱收敛于某一有界的非降函数 $F^*(x)$，这个极限函数至少在 x_0 点与 $F(x)$ 不相等。但重复前面的论证可知 $F^*(x)$ 亦应是分布函数，其对应的特征函数也是 $f(t)$，由唯一性定理，我们又有 $F^*(x)=F(x)$，引出了矛盾。故 $\{F_n(x)\}$ 弱收敛于 $F(x)$，于是证得定理。

在逆极限定理中，若保留“特征函数列 $\{f_n(t)\}$ 收敛于某一函数 $f(t)$”的要求，而把“$f(t)$ 在 $t=0$ 连续”的要求改成“特征函数列 $\{f_n(t)\}$ 在包含原点的某一区间中一致收敛于函数 $f(t)$”，则定理的结论仍然成立。这是因为由一致收敛性及 $f_n(t)$ 在原点的连续性可以推知 $f(t)$ 在原点的连续性。通常把“特征函数列 $\{f_n(t)\}$ 在 $(-\infty,\ \infty)$ 上的任一有限闭区间中都一致收敛于一个函数 $f(t)$”简称为“$\{f_n(t)\}$ 内闭匀敛于 $f(t)$”，这样我们就可以把分布函数列 $\{F_n(x)\}$ 弱收敛于某一分布函数的充要条件简述为：它相应的特征函数列 $\{f_n(t)\}$ 内闭匀敛于某一函数 $f(t)$。

通常把正逆极限定理合称**连续性定理**，因为它们表述了分布函数与特征函数一一对应关系的“连续性”。这定理最先由法国数学家勒维(Lévy)及瑞典数学家克拉美(Cramér)证得，因此又称**勒维-克拉美定理**。

三、随机变量的收敛性

概率论中的极限定理研究的是随机变量序列的某种收敛性，对随机变量收敛性的不同定义将导致不同的极限定理，而随机变量的收敛性的确可以有各种不同的定义，现在就来讨论这个问题。

首先，分布函数弱收敛的讨论启发我们引进如

定义 2（依分布收敛） 设随机变量 $\xi_n(\omega)$、$\xi(\omega)$

分别为 $F_n(x)$ 及 $F(x)$，如果 $F_n(x)\xrightarrow{W}F(\omega)$，则称$\{\xi_n(\omega)\}$**依分布收敛于** $\xi(\omega)$，并记为 $\xi_n(\omega)\xrightarrow{L}\xi(\omega)$.

其次，由贝努里大数定律，我们很自然地引进下面的定义。

定义 3（依概率收敛） 如果

$$\lim_{n\to\infty}P\{|\xi_n(\omega)-\xi(\omega)|\geqslant\varepsilon\}=0 \tag{19}$$

对任意的 $\varepsilon>0$ 成立，则称$\{\xi_n(\omega)\}$**依概率收敛于** $\xi(\omega)$，并记为 $\xi_n(\omega)\xrightarrow{P}\xi(\omega)$.

这样一来，贝努里大数定律可以重新叙述如下：

设 μ_n 是 n 次独立试验中事件 A 出现的次数，而 p 是事件 A 在每次试验中出现的概率，则频率$\frac{\mu_n}{n}$依概率收敛于概率 p.

上述两种收敛性之间的关系可以从下面定理中看到，这也说明了随机变量序列依概率收敛性的重要性。

定理 6 $\xi_n\xrightarrow{P}\xi\Longrightarrow\xi_n\xrightarrow{L}\xi$.

[证明] 因为，对 $x'<x$ 有

$$\begin{aligned}\{\xi<x'\}&=\{\xi_n<x,\xi<x'\}+\{\xi_n\geqslant x,\xi<x'\}\\&\subset\{\xi_n<x\}+\{\xi_n\geqslant x,\xi<x'\}\end{aligned}$$

所以我们有

$$F(x')\leqslant F_n(x)+P\{\xi_n\geqslant x,\xi<x'\}$$

如果$\{\xi_n\}$依概率收敛于 ξ，则

$$P\{\xi_n\geqslant x,\xi<x'\}\leqslant P\{|\xi_n-\xi|\geqslant x-x'\}\to0$$

因而有

$$F(x')\leqslant\varliminf_{n\to\infty}F_n(x)$$

同理可证，对 $x''>x$，成立

$$\overline{\lim_{n\to\infty}}F_n(x)\leqslant F(x'')$$

所以对 $x'<x<x''$，有

$$F(x')\leqslant\underline{\lim_{n\to\infty}}F_n(x)\leqslant\overline{\lim_{n\to\infty}}F_n(x)\leqslant F(x'')$$

如果 x 是 $F(x)$ 的连续点，则令 x', x'' 趋于 x 可得

$$F(x)=\lim_{n\to\infty}F_n(x)$$

定理证毕.

由于不同的随机变量可以对应于同一分布函数，因此一般地讲，由分布函数列的收敛性当然推不出随机变量序列的其它收敛性. 试看下例.

[例 3] 若样本空间 $\Omega=\{\omega_1,\omega_2\}$，$P(\omega_1)=P(\omega_2)=\frac{1}{2}$，定义随机变量 $\xi(\omega)$ 如下：$\xi(\omega_1)=-1$，$\xi(\omega_2)=1$，则 $\xi(\omega)$ 的分布列为

$$\begin{pmatrix}-1, & 1\\ \frac{1}{2}, & \frac{1}{2}\end{pmatrix} \tag{20}$$

若对一切 n，令 $\xi_n(\omega)=-\xi(\omega)$，显然 $\xi_n(\omega)$ 的分布列也是(20)，因此 $\xi_n(\omega)\xrightarrow{L}\xi(\omega)$. 但是对任意的 $0<\varepsilon<2$，

$$P\{|\xi_n(\omega)-\xi(\omega)|>\varepsilon\}=P\{\Omega\}=1$$

因此 $\{\xi_n(\omega)\}$ 不依概率收敛于 $\xi(\omega)$，这为定理 6 之逆提供了反例.

进一步，若令 $\xi_{2n}(\omega)=\xi(\omega)$，$\xi_{2n+1}(\omega)=-\xi(\omega)$ 则 $\xi_n(\omega)\xrightarrow{L}\xi(\omega)$ 依然成立，但此时对随机变量序列 $\{\xi_n(\omega)\}$ 实在很难有其他的收敛性可言.

但是，在特殊场合却有下面结果.

定理 7 设 C 是常数，则 $\xi_n\xrightarrow{P}C\Longleftrightarrow\xi_n\xrightarrow{L}C$.

［证明］ 由定理6可知只须证明由依分布收敛于常数可推出依概率收敛于常数．事实上，对任意的 $\varepsilon>0$，

$$P\{|\xi_n-C|\geqslant\varepsilon\}=P\{\xi_n\geqslant C+\varepsilon\}+P\{\xi_n\leqslant C-\varepsilon\}$$
$$=1-F_n(C+\varepsilon)+F_n(C-\varepsilon+0)\to1-1+0=0$$
$$(n\to\infty)$$

因为有的大数定律是讨论随机变量序列收敛于常数的，这时将用到上述结果．

仔细考察上节关于大数定律的证明，有助于理解下面关于随机变量收敛性的第三种定义．

定义4（r-阶收敛） 设对随机变量 ξ_n 及 ξ 有 $E|\xi_n|^r<\infty$，$E|\xi|^r<\infty$，其中 $r>0$ 为常数，如果

$$\lim_{n\to\infty}E|\xi_n-\xi|^r=0 \tag{21}$$

则称$\{\xi_n\}$ r-阶收敛于 ξ，并记为 $\xi_n\xrightarrow{r}\xi$．

下面定理揭示了 r-阶收敛与依概率收敛的关系．

定理8 $\xi_n\xrightarrow{r}\xi\Longrightarrow\xi_n\xrightarrow{P}\xi$．

［证明］ 先证对于任意 $\varepsilon>0$，成立

$$P\{|\xi_n-\xi|\geqslant\varepsilon\}\leqslant\frac{E|\xi_n-\xi|^r}{\varepsilon^r} \tag{22}$$

事实上，若以 $F(x)$ 记 $\xi_n-\xi$ 的分布函数，则仿车贝晓夫不等式的证明可得

$$P\{|\xi_n-\xi|\geqslant\varepsilon\}=\int_{|x|\geqslant\varepsilon}dF(x)$$
$$\leqslant\int_{|x|\geqslant\varepsilon}\frac{|x|^r}{\varepsilon^r}dF(x)\leqslant\frac{1}{\varepsilon^r}\int_{-\infty}^{\infty}|x|^r dF(x)$$
$$=\frac{E|\xi_n-\xi|^r}{\varepsilon^r}$$

不等式(22)是车贝晓夫不等式的推广，通常称作马尔可夫不等式，当 $r=2$ 时就是车贝晓夫不等式．定理8是马尔可夫不等式

的直接推论.

下例说明定理 8 之逆不真.

[例 4] 取 $\Omega=(0,1]$, $\mathscr{F}$为(0, 1]中波雷尔点集全体所构成的σ-域, P 为勒贝格测度. 定义 $\xi(\omega)\equiv 0$ 及

$$\xi_n(\omega)=\begin{cases} n^{1/r}, & 0<\omega\leqslant\dfrac{1}{n} \\ 0, & \dfrac{1}{n}<\omega\leqslant 1 \end{cases} \tag{23}$$

显然对一切 $\omega\in\Omega$, $\xi_n(\omega)\longrightarrow\xi(\omega)$, 又对于任意的 $\varepsilon>0$,

$$P\{|\xi_n(\omega)-\xi(\omega)|\geqslant\varepsilon\}\leqslant\frac{1}{n}$$

因此 $\xi_n\xrightarrow{P}\xi$, 但是

$$E|\xi_n-\xi|^r=(n^{1/r})^r\cdot\frac{1}{n}=1$$

在 r-阶收敛中, 最重要的是 $r=2$ 的情况, 这时称为**均方收敛**, 在第三部分第三章 § 1 中, 对均方收敛有深入的讨论.

下面是关于随机变量收敛性的第四种定义.

定义 5 (概率 1 收敛) 如果

$$P\{\lim_{n\to\infty}\xi_n(\omega)=\xi(\omega)\}=1 \tag{24}$$

则称$\{\xi_n(\omega)\}$以概率 1 收敛于$\xi(\omega)$, 又称$\{\xi_n(\omega)\}$**几乎处处收敛于**$\xi(\omega)$, 记为 $\xi_n(\omega)\xrightarrow{a.s.}\xi(\omega)$.

概率 1 收敛是概率论中较强的一种收敛性, 但是正如例 4 所表明的, 一般并不能由它推出 r-阶收敛. 关于概率 1 收敛的讨论将在 § 4 中继续进行, 在那里将证明可由概率 1 收敛推出依概率收敛.

*四、波赫纳尔(Bochner)-辛钦(Хинчин)定理

利用这个机会, 我们来叙述并证明一个关于特征函数的重要

定理.

定理 9（波赫纳尔-辛钦） 函数 $f(t)$ 是特征函数的充要条件是：$f(t)$ 非负定，连续，且 $f(0)=1$.

在证明的过程中，顺带证明了一个在随机过程中将用到的与上述定理类似的赫尔格洛兹(Herglotz)定理．为此，有

定义 6 如果对任意的正整数 n 及复数 $\lambda_1, \lambda_2, \cdots, \lambda_n$ 均有

$$\sum_{k=1}^{n}\sum_{j=1}^{n}C_{k-j}\lambda_k\bar{\lambda}_j\geqslant 0 \tag{25}$$

则称复数列 $C_n(n=0, \pm 1, \pm 2, \cdots)$ 是非负定的.

定理 10（赫尔格洛兹） 数列 $C_n(n=0, \pm 1, \pm 2, \cdots)$ 可以表为

$$C_n=\int_{-\pi}^{\pi}e^{inx}dG(x) \tag{26}$$

的充要条件是它是非负定的，其中 $G(x)$ 是 $[-\pi, \pi]$ 上有界、非降、左连续函数.

定理 9 的必要性已在第四章 § 4 中证过，定理 10 也可类似证明，下面只须证明充分性.

由于 $f(t)$ 是非负定的，故对任何 N，实数 $\frac{k}{n}$ 及复数 e^{-ikx} $(k=0, 1, \cdots, N-1)$，皆有

$$\mathscr{P}_N^{(n)}(x)=\frac{1}{N}\sum_{k=0}^{N-1}\sum_{j=0}^{N-1}f\left(\frac{k-j}{n}\right)e^{-i(k-j)x}\geqslant 0$$

易知，其中使 $k-j$ 等于 r 的项有 $N-|r|$ 个，r 可由 $-N+1$ 变到 $N-1$．因此

$$\mathscr{P}_N^{(n)}(x)=\sum_{r=-N}^{N}\left(1-\frac{|r|}{N}\right)f\left(\frac{r}{n}\right)e^{-irx}$$

从而

$$\int_{-\pi}^{\pi} e^{isx}\mathscr{P}_N^{(n)}(x)dx$$

$$=\sum_{r=-N}^{N}\left(1-\frac{|r|}{N}\right)f\left(\frac{r}{n}\right)\int_{-\pi}^{\pi} e^{-irx}e^{isx}dx$$

由于

$$\int_{-\pi}^{\pi} e^{-i(r-s)x}dx=\begin{cases}0, & r\neq s\\ 2\pi, & r=s\end{cases}$$

所以

$$\left(1-\frac{|s|}{N}\right)f\left(\frac{s}{n}\right)=\frac{1}{2\pi}\int_{-\pi}^{\pi} e^{isx}\mathscr{P}_N^{(n)}(x)dx$$

$$=\int_{-\pi}^{\pi} e^{isx}dF_N^{(n)}(x)$$

其中

$$F_N^{(n)}(x)=\frac{1}{2\pi}\int_{-\pi}^{x}\mathscr{P}_N^{(n)}(t)dt$$

是一个在$[-\pi,\pi]$上有界的非降函数，其全变差为

$$F_N^{(n)}(\pi)=\frac{1}{2\pi}\int_{-\pi}^{\pi}\mathscr{P}_N^{(n)}(t)dt=f(0)=1$$

补充定义 $x<-\pi$ 时 $F_N^{(n)}(x)=0$，$x>\pi$ 时，$F_N^{(n)}(x)=1$，则 $F_N^{(n)}(x)$ 是一分布函数.

按海莱第一定理，存在序列 N_k，$k\to\infty$ 时 $N_k\to\infty$，使函数序列 $F_{N_k}^{(n)}(x)$弱收敛于某一非降函数 $F^{(n)}(x)$. 又因对任何N及 $e>0$,

$$F_N^{(n)}(-\pi-e)=0,\quad F_N^{(n)}(\pi+e)=1$$

因而也有

$$F^{(n)}(-\pi-e)=0,\quad F^{(n)}(\pi+e)=1$$

所以 $F^{(n)}(x)$也是分布函数.

按海莱第二定理，

$$\lim_{k\to\infty}\int_{-\pi}^{\pi} e^{isx}dF_{N_k}^{(n)}(x)=\int_{-\pi}^{\pi} e^{isx}dF^{(n)}(x)$$

所以，对一切整数 $s(s=0, \pm1, \pm2, \cdots)$有

$$f\left(\frac{s}{n}\right)=\int_{-\pi}^{\pi} e^{isx}dF^{(n)}(x)$$

至此我们已顺便证明了赫尔格洛兹定理．特别地

$$f\left(\frac{1}{n}\right)=\int_{-\pi}^{\pi} e^{ix}dF^{(n)}(x) \tag{27}$$

考虑特征函数序列

$$f_n(t)=\int_{-n\pi}^{n\pi} e^{itx}dF_n(x)$$

其中 $F_n(x)=F^{(n)}\left(\frac{x}{n}\right)$．易知对一切整数 k 有

$$f_n\left(\frac{k}{n}\right)=f\left(\frac{k}{n}\right) \tag{28}$$

对任何 t，我们总能选取序列 $k=k(n, t)$，使 $0\leqslant t-\frac{k}{n}<\frac{1}{n}$．由于 $f(t)$ 连续，从而

$$f(t)=\lim_{n\to\infty}f\left(\frac{k}{n}\right)=\lim_{n\to\infty}f_n\left(\frac{k}{n}\right) \tag{29}$$

如能证明对一切实数 t，有

$$f(t)=\lim_{n\to\infty}f_n(t) \tag{30}$$

那么由逆极限定理即知 $f(t)$是特征函数了．

为此，由(28)和(29)有

$$\begin{aligned}\lim_{n\to\infty}f_n(t)&=\lim_{n\to\infty}\left\{\left[f_n(t)-f_n\left(\frac{k}{n}\right)\right]+f_n\left(\frac{k}{n}\right)\right\}\\&=f(t)+\lim_{n\to\infty}\left[f_n(t)-f_n\left(\frac{k}{n}\right)\right]\end{aligned} \tag{31}$$

令 $\theta=t-\frac{k}{n}$，那末 $0\leqslant\theta<\frac{1}{n}$．按 $f_n(t)$的定义，有

$$\left|f_n(t)-f_n\left(\frac{k}{n}\right)\right|=\left|\int_{-n\pi}^{n\pi} e^{i(k/n)x}(e^{i\theta x}-1)dF_n(x)\right|$$

$$\leqslant \int_{-n\pi}^{n\pi} |e^{i\theta x}-1| dF_n(x) \tag{32}$$

利用许瓦兹不等式，可得

$$\int_{-n\pi}^{n\pi} |e^{i\theta x}-1| dF_n(x) \leqslant \sqrt{\int_{-n\pi}^{n\pi} |e^{i\theta x}-1|^2 dF_n(x)}$$

$$=\left[\int_{-n\pi}^{n\pi} 2(1-\cos\theta x) dF_n(x)\right]^{\frac{1}{2}}$$

$$=[2(1-\mathrm{Re}f_n(\theta))]^{\frac{1}{2}} \tag{33}$$

其中 $\mathrm{Re}f_n(\theta)$ 为 $f_n(\theta)$ 的实数部分，既然在 $0\leqslant\alpha<1$ 及 $-\pi\leqslant z<\pi$ 时有 $\cos z\leqslant\cos\alpha z$，则

$$1-\mathrm{Re}f_n(\theta)=\int_{-n\pi}^{n\pi}(1-\cos\theta x)dF_n(x)$$

$$=\int_{-\pi}^{\pi}(1-\cos\theta nz)dF_n(nz)\leqslant\int_{-\pi}^{\pi}(1-\cos z)dF_n(nz)$$

$$=\int_{-\pi}^{\pi}(1-\cos z)dF^{(n)}(z)$$

$$=1-\mathrm{Re}\int_{-\pi}^{\pi}e^{iz}dF^{(n)}(z)$$

再由(27)我们得到

$$1-\mathrm{Re}f_n(\theta)\leqslant 1-\mathrm{Re}f\left(\frac{1}{n}\right) \tag{34}$$

合并(31)，(32)，(33)，即得

$$\left|f_n(t)-f_n\left(\frac{k}{n}\right)\right|\leqslant\sqrt{2\left(1-\mathrm{Re}f\left(\frac{1}{n}\right)\right)}$$

注意到 $f(0)=1$，则由 $f(t)$ 的连续性推得

$$\lim_{n\to\infty}\left[f_n(t)-f_n\left(\frac{k}{n}\right)\right]=0$$

于是由(31)知(30)成立，定理证毕.

若不证赫尔格洛兹定理而直接证波赫纳尔-辛钦定理，有比较简炼的证法，可参看 C.R.Rao, *Linear Statistical Inference*

and its Applications, second edition, pp. 141—142.

§3. 独立同分布场合的极限定理

一、独立和问题

在§1中，我们讨论了贝努里试验场合事件A出现次数μ_n的极限行为，曾指出μ_n可以表示为n个独立随机变量之和(以后简称"独立和")，并对它证明了大数定律及中心极限定理．后来又看到这些定理有重要应用．这里自然会提出这样一个问题：这些性质是否只在贝努里试验场合才具有？

研究表明，许多独立和具有类似的性质，本节就要进一步讨论这个问题．

独立和的问题经常出现，例如测量一物体的某种尺寸，譬如测量一个圆柱体的直径d，通常采用的办法是对它进行n次测量，得到数值$\xi_1, \xi_2, \cdots, \xi_n$，然后采用平均值

$$\eta_n=\frac{\xi_1+\xi_2+\cdots+\xi_n}{n}$$

作为d的数值．我们知道，测量时有各种随机因素影响，因此其结果带有随机性，这时η_n是随机变量之和，如果各次测量是独立的，η_n便是独立和．为了说明上面所用办法的合理性就必须研究独立和．但这里的ξ_n不服从贝努里0—1分布，因此已不是贝努里试验场合的问题了．

在数理统计中已经把上述做法一般化．为了研究母体(它通常描述我们感兴趣的某一类现象)的某些特征，就对母体进行若干次观察以得到一批观察值$\xi_1, \xi_2, \cdots, \xi_n$，并称它们是一个容量为$n$的子样．再利用这个子样来构造各种统计量，例如

$\frac{\xi_1+\xi_2+\cdots+\xi_n}{n}$或$\frac{\xi_1^2+\xi_2^2+\cdots+\xi_n^2}{n}$，以对母体的特征作各种推

断．虽然每次观察得到的是具体的数值，但是为了比较各个统计量或各种推断方法的优劣，有必要把这些观察看作是某随机变量 ξ 的观察值，通常假定 $\xi_1, \xi_2, \cdots, \xi_n$ 是相互独立的，且它们与 ξ 具有相同的概率分布，这时上面的两个统计量便都是独立同分布的随机变量之和.

独立和的问题在许多实际问题中也出现，例如在计算电车整流站的电力负荷时，就遇到独立和问题，因为整流站的电力负荷等于各电车使用电力之和，每辆电车在某时刻的用电量是随机的，作为初步近似，可以假定各电车的用电量是相互独立的，因此这里遇到的正是独立和．在车间用电问题中，若有多类车床，用电量各不相同，则总用电量也需通过独立和来计算.

可见独立和问题经常遇到，而且各加项一般都不是0—1分布．本节专门讨论各个加项服从相同分布的场合，这是实际工作(特别是在数理统计)中最常碰到的．从数学方面来看，这是最简单、最基本、最便于处理的，而且所用的处理方法可以相当方便地运用到更一般的场合.

在上章§4中，我们讨论过独立的服从同类型分布的随机变量之和的分布问题，证明了某些分布的再生性，使用的工具是特征函数．现在我们要处理的问题与那里有很大区别，首先我们将对很一般的分布进行讨论，因而再生性通常都不满足；其次，不是对固定的 n 进行讨论，而是讨论 $n \to \infty$ 时的情况，即研究极限定理．从数学的角度来看，它们可以看作是贝努里试验场合极限定理的推广，这里也研究大数定律与中心极限定理.

所使用的工具还是特征函数．我们已经看到它很适合于处理独立和问题，有了上节的连续性定理，我们将进一步看到，它也很适合于处理极限分布问题．事实上，正是由于特征函数这一有力工具的使用，使得所有古典极限定理在短期内便得到了完满的解

决．拉普拉斯就已经知道并应用了特征函数，俄国数学家李雅普洛夫(Ляпунов)最先发现并证明了收敛于正态分布的连续性定理(但并未明确叙述)，从那时起，特征函数的理论不断得到完善．在这当中，法国数学家勒维有突出的贡献．现在，特征函数法已经成了概率论的基本方法之一．

二、辛钦大数定律

在§1中，我们已经通过车贝晓夫不等式建立起多种大数定律，那里都假定了方差的存在性，但是在独立同分布场合，并不需要有这个要求，这就是有名的辛钦大数定律告诉我们的．用特征函数作为工具，这个定理很容易证明．

定理1(辛钦)　设 $\xi_1, \xi_2, \cdots, \xi_n, \cdots$ 是相互独立的随机变量序列，它们服从相同的分布，且具有有限的数学期望

$$a = E\xi_n$$

则对任意的 $\varepsilon>0$，有

$$\lim_{n\to\infty} P\left\{\left|\frac{1}{n}\sum_{i=1}^{n}\xi_i - a\right| < \varepsilon\right\} = 1 \tag{1}$$

［证明］　由于 $\xi_1, \xi_2, \cdots, \xi_n$ 具有相同分布，故有同一特征函数，设为 $f(t)$，因为数学期望存在，故 $f(t)$ 可展开成

$$f(t) = f(0) + f'(0)t + o(t) = 1 + iat + o(t) \tag{2}$$

而 $\frac{1}{n}\sum_{i=1}^{n}\xi_i$ 的特征函数为

$$\left[f\left(\frac{t}{n}\right)\right]^n = \left[1 + ia\frac{t}{n} + o\left(\frac{t}{n}\right)\right]^n \tag{3}$$

对于固定的 t

$$\left[f\left(\frac{t}{n}\right)\right]^n \longrightarrow e^{iat} \quad (n\to\infty) \tag{4}$$

极限函数 e^{iat} 是连续函数，它是退化分布 $I(x-a)$ 所对应的特征

函数，由逆极限定理知$\frac{1}{n}\sum_{i=1}^{n}\xi_i$的分布函数弱收敛于$I(x-a)$，再由上节定理7知$\frac{1}{n}\sum_{i=1}^{n}\xi_i$依概率收敛于常数$a$，从而证明了定理.

显然，贝努里大数定律是辛钦大数定律的特殊情况．辛钦大数定律在应用中很重要，下面通过一个例子来说明.

［例1］（用蒙特卡洛方法计算定积分）为计算积分

$$J=\int_a^b g(x)dx \tag{5}$$

可以通过下面概率论方法实现.

任取一列相互独立的、都具有$[a,\ b]$中均匀分布的随机变量$\{\xi_i\}$，则$\{g(\xi_i)\}$也是一列相互独立相同分布的随机变量，而且

$$Eg(\xi_i)=\frac{1}{b-a}\int_a^b g(x)dx=\frac{J}{b-a} \tag{6}$$

既然

$$J=(b-a)\cdot Eg(\xi_i) \tag{7}$$

因此只要能求得$Eg(\xi_i)$，便能得到J的数值.

为求$Eg(\xi_i)$，自然想到大数定律，因为

$$\frac{g(\xi_1)+g(\xi_2)+\cdots+g(\xi_n)}{n}\xrightarrow{P}Eg(\xi_i) \tag{8}$$

这样一来，只要能生成随机变量序列$\{g(\xi_i)\}$就能对积分(5)进行数值计算，而生成$\{g(\xi_i)\}$的关键是要生成相互独立相同分布的$\{\xi_i\}$，这里的ξ_i均服从$[a,b]$上均匀分布.

现在已经可以把上述想法变成现实．这就是在电子计算机上产生服从均匀分布$[a,b]$的随机数$\{\xi_i\}$并通过(7)(8)算出J，这种做法与我们在蒲丰投针问题中通过投针计算圆周率π的做法是一致的．这种通过概率论的想法构造模型从而实现数值计算的方法，随着电子计算机的发展，已形成一种新的计算方法——**概率计算**

方法，亦称蒙特卡洛方法，它在原子物理、公用事业理论中发挥了不少作用，这个方法的理论根据之一就是大数定律.

至于计算积分，蒙特卡洛方法的实用场合是计算重积分

$$I=\int_K g(P)dP \tag{9}$$

其中P是m维空间的点，当m较大时，用蒙特卡洛方法比一般数值法有优点，主要是它的误差与维数m无关.

对于蒙特卡洛方法的进一步讨论在第三部分第六章专门进行.

三、中心极限定理

我们转而考虑如何把积分极限定理推广到相互独立相同分布，但分布函数为任意的随机变量序列的场合，这类问题在实际应用中非常重要.

若$\xi_1, \xi_2, \cdots, \xi_n, \cdots$是一串相互独立相同分布的随机变量序列，且

$$E\xi_k=m,\quad D\xi_k=\sigma^2 \tag{10}$$

我们来讨论标准化随机变量和

$$\zeta_n=\frac{1}{\sigma\sqrt{n}}\sum_{k=1}^{n}(\xi_k-m) \tag{11}$$

的极限分布.

林德贝格(Lindeberg)与勒维建立了下列中心极限定理.

定理2(林德贝格-勒维)　若$0<\sigma^2<\infty$，则

$$\lim_{n\to\infty}P\{\zeta_n<x\}=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-t^2/2}dt \tag{12}$$

[证明]　记ξ_k-m的特征函数为$\varphi(t)$，则ζ_n的特征函数为$\left[\varphi\left(\frac{t}{\sigma\sqrt{n}}\right)\right]^n$. 由于$E\xi_k=m$，$D\xi_k=\sigma^2$故$\varphi'(0)=0$，$\varphi''(0)=-\sigma^2$. 因此

$$\varphi(t)=1-\frac{1}{2}\sigma^2t^2+o(t^2) \tag{13}$$

所以

$$\left[\varphi\left(\frac{t}{\sigma\sqrt{n}}\right)\right]^n=\left[1-\frac{1}{2n}t^2+o\left(\frac{t^2}{n}\right)\right]^n\longrightarrow e^{-t^2/2} \tag{14}$$

由于 $e^{-t^2/2}$ 是连续函数，它对应的分布函数为 $N(0, 1)$，因此由逆极限定理知

$$P\{\zeta_n<x\}\longrightarrow\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-t^2/2}dt$$

定理证毕.

用这个定理立即可以推出德莫哇佛-拉普拉斯积分极限定理.

林德贝格-勒维定理有广泛应用. 在实际工作中，只要 n 足够大，便可以把独立同分布的随机变量之和当作是正态变量. 这种做法在数理统计中用得很普遍，当处理大子样(容量很大的子样)时，林德贝格-勒维定理是重要工具.

下面我们介绍另外的两个例子.

[例 2] (正态随机数的产生)在蒙特卡洛方法中经常需要产生服从正态分布的随机数，但是一般计算机只备有产生[0, 1]均匀分布随机数(实际上是伪随机数)的程序. 怎样通过[0, 1]均匀分布的随机数来产生正态随机数呢？这有多种途径，最常用的是利用上述定理来实现.

设 $\xi_1, \xi_2, \cdots, \xi_n, \cdots$ 是相互独立、均服从[0, 1]均匀分布的随机变量，这时定理 2 的条件得到满足，故 $\xi_1+\xi_2+\cdots+\xi_n$ 渐近于正态变量. 一般 n 取不太大的值就可满足实际要求. 图 1 中给出了 $n=1, 2, 3$ 时的图象. 在蒙特卡洛中，一般取 $n=12$，并用下式得到新的随机数序列

$$\eta_k=\sum_{i=1}^{12}\xi_{12(k-1)+i}-6,\quad k=1, 2, \cdots \tag{15}$$

显然$\{\eta_k\}$也是独立随机数序列，而且 $E\eta_k=0$，$D\eta_k=1$. 经过检验证明，这时 η_k 的渐近正态性已能满足一般精度要求.

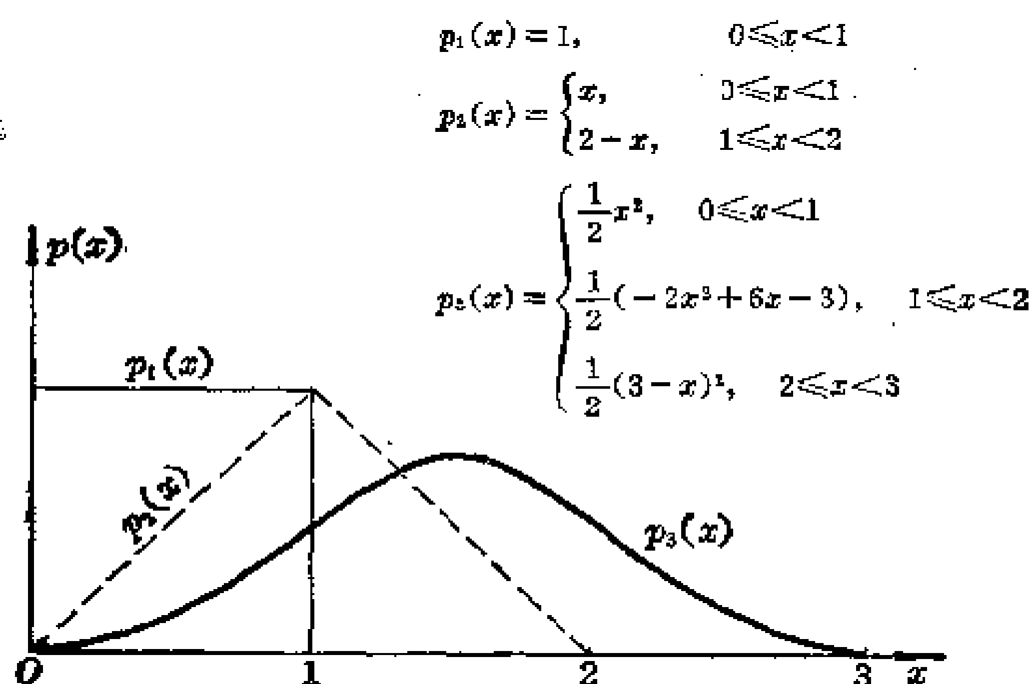


图 1　均匀分布褶积

［例 3］（近似数定点运算的误差分析）数值计算时，任何数 x 都只能用一定位数的有限小数 y 来近似，这就产生了一个误差 $\xi=x-y$，在下面讨论中，我们假定参加运算的数都用十进制定点表示，每个数都用四舍五入的方法取到小数点后五位，这时相应的舍入误差可以看作是$[-0.5\times10^{-5},\ 0.5\times10^{-5})$上的均匀分布.

现在如果要求 n 个数 $x_i(i=1,2,\cdots,n)$的和 S，在数值计算中就只能求出相应的有限位小数 $y_i(i=1,2,\cdots,n)$的和 T，并用 T 作为 S 的近似值. 自然要问，这样做造成的误差 $\eta=S-T$ 是多少？

因为我们有

$$S=\sum_{i=1}^{n}x_i=\sum_{i=1}^{n}(y_i+\xi_i)=\sum_{i=1}^{n}y_i+\sum_{i=1}^{n}\xi_i$$

故

$$\eta=\sum_{i=1}^{n}\xi_i$$

一种传统的估计方法是这样的：由于

$$|\xi_i| \leqslant 0.5 \times 10^{-5}$$

所以

$$|\eta| \leqslant \sum_{i=1}^{n} |\xi_i| \leqslant n \times 0.5 \times 10^{-5}$$

以 $n=10000$ 为例，所得的误差估计为

$$|\eta| \leqslant 0.05 \tag{16}$$

这种估计方法显然太保守，看来用概率论方法估计是适宜的。这时直接求 $\eta = \sum_{i=1}^{n} \xi_i$ 的分布不容易，但当 n 较大时用极限定理作为工具，则能使问题很快得到解决．因为

$$m = E\xi_i = 0, \quad \sigma = \sqrt{D\xi_i} = \frac{0.5 \times 10^{-5}}{\sqrt{3}}$$

如果假定舍入误差 ξ_i 是相互独立的，n 又较大，那么用定理 2 得到

$$P\left\{\left|\sum_{i=1}^{n} \xi_i\right| < k\sqrt{n}\,\sigma\right\} \approx \frac{1}{\sqrt{2\pi}} \int_{-k}^{k} e^{-t^2/2} dt$$

取 $k=3$ 时，上式右边为 0.997，因此我们能以 99.7% 的概率断言：

$$|\eta| < 3 \times 100 \times \frac{0.5 \times 10^{-5}}{\sqrt{3}} = 0.866 \times 10^{-3} \tag{17}$$

这只及(16)式中上限估计的 60 分之一．

历史上，误差分析是概率论的重要生长点之一．19 世纪初德国数学家高斯(Gauss) 正是在研究测量误差时引进了正态分布并发展了有广泛应用的最小二乘法．至今这仍是概率论与生产实际有广泛联系的领域之一．由于高斯对正态分布的研究，因此有不少人称正态分布为**高斯分布**（虽然德莫哇佛及拉普拉斯早在高斯之前就把正态分布引入概率论）．

*§4. 强大数定律

一、概率1收敛

以前，我们曾顺带地提起过概率1收敛及强大数定律，本节将对它们进行深入讨论.

要彻底搞清概率1收敛这个概念，必须对事件(点集)序列的运算有进一步了解，我们就从讨论这个问题开始.

设 $A_1, A_2, \cdots, A_n, \cdots$ 是一列事件，则 $\bigcup_{n=k}^{\infty} A_n$ 表示事件序列 A_k, $A_{k+1}, \cdots$ 中至少发生一个，而 $\bigcap_{n=k}^{\infty} A_n$ 则表示 A_k, A_{k+1}, $\cdots$ 同时发生。记

$$\overline{\lim_{n\to\infty}} A_n = \bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} A_n \tag{1}$$

$$\underline{\lim_{n\to\infty}} A_n = \bigcup_{k=1}^{\infty} \bigcap_{n=k}^{\infty} A_n \tag{2}$$

称 $\overline{\lim\limits_{n\to\infty}} A_n$ 为事件序列 $\{A_n\}$ 的**上限事件**，它表示 A_n 发生无穷多次，因为 $\omega \in \bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} A_n$ 当且仅当 ω 属于无穷多个 A_n；类似地称 $\underline{\lim\limits_{n\to\infty}} A_n$ 为事件序列 $\{A_n\}$ 的**下限事件**，它表示 A_n 至多只有有限个不发生，因为 $\omega \in \bigcup_{k=1}^{\infty} \bigcap_{n=k}^{\infty} A_n$，当且仅当存在一个 N，使 $\omega \in \bigcap_{n=N}^{\infty} A_n$，因此若 ω 发生，则 $A_N, A_{N+1}, \cdots$ 同时发生，这时至多只有前面 $N-1$ 个事件 $A_1, A_2, \cdots, A_{N-1}$ 可能不发生(也可能发生)。

显然

$$\overline{\lim_{n\to\infty}} A_n \supset \underline{\lim_{n\to\infty}} A_n \tag{3}$$

特别当$\overline{\lim\limits_{n\to\infty}}A_n=\underline{\lim\limits_{n\to\infty}}A_n$时，记$\lim\limits_{n\to\infty}A_n\equiv\overline{\lim\limits_{n\to\infty}}A_n=\underline{\lim\limits_{n\to\infty}}A_n$，并称它为事件序列$\{A_n\}$的**极限事件**.

利用德莫根定理，有

$$\overline{\left(\bigcap_{k=1}^{\infty}\bigcup_{n=k}^{\infty}A_n\right)}=\bigcup_{k=1}^{\infty}\bigcap_{n=k}^{\infty}\overline{A}_n$$

$$\overline{\left(\bigcup_{k=1}^{\infty}\bigcap_{n=k}^{\infty}A_n\right)}=\bigcap_{k=1}^{\infty}\bigcup_{n=k}^{\infty}\overline{A}_n$$

因此

$$\underline{\lim_{n\to\infty}}\overline{A}_n=\overline{\left(\overline{\lim_{n\to\infty}}A_n\right)} \tag{4}$$

$$\overline{\lim_{n\to\infty}}\overline{A}_n=\overline{\left(\underline{\lim_{n\to\infty}}A_n\right)} \tag{5}$$

下面波雷尔-康特立（Cantelli）引理在概率论中有众多的应用.

引理1（波雷尔-康特立引理）

(i) 若随机事件序列$\{A_n\}$满足

$$\sum_{n=1}^{\infty}P(A_n)<\infty \tag{6}$$

则

$$P\{\overline{\lim_{n\to\infty}}A_n\}=0,\qquad P\{\underline{\lim_{n\to\infty}}\overline{A}_n\}=1 \tag{7}$$

(ii) 若$\{A_n\}$是相互独立的随机事件序列，则

$$\sum_{n=1}^{\infty}P(A_n)=\infty \tag{8}$$

成立的充要条件为

$$P\{\overline{\lim_{n\to\infty}}A_n\}=1\quad 或\quad P\{\underline{\lim_{n\to\infty}}\overline{A}_n\}=0 \tag{9}$$

[证明] (i) 由于

$$P\{\overline{\lim_{n\to\infty}} A_n\} = P\left\{\bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} A_n\right\}$$

$$\leqslant P\left\{\bigcup_{n=k}^{\infty} A_n\right\} \leqslant \sum_{n=k}^{\infty} P\{A_n\} \to 0 \qquad (k\to\infty)$$

由(4)

$$P\{\underline{\lim_{n\to\infty}} \bar{A}_n\} = 1$$

(ii) 先证必要性. 注意到$\{A_n\}$的独立性, 有

$$P\{\underline{\lim_{n\to\infty}} \bar{A}_n\} = P\left\{\bigcup_{k=1}^{\infty} \bigcap_{n=k}^{\infty} \bar{A}_n\right\} \leqslant \sum_{k=1}^{\infty} P\left\{\bigcap_{n=k}^{\infty} \bar{A}_n\right\}$$

$$= \sum_{k=1}^{\infty} \prod_{n=k}^{\infty} P(\bar{A}_n) = \sum_{k=1}^{\infty} \prod_{n=k}^{\infty} [1 - P(A_n)] \tag{10}$$

因$\sum_{n=k}^{\infty} P(A_n) = \infty$, 从无穷乘积理论可知$\prod_{n=k}^{\infty} [1 - P(A_n)] = 0$, 所以

$$P\{\underline{\lim_{n\to\infty}} \bar{A}_n\} = 0$$

再证充分性. 若 $P\{\overline{\lim_{n\to\infty}} A_n\} = 1$. 假定$\sum_{n=1}^{\infty} P(A_n) < \infty$, 则由(i)得到 $P\{\overline{\lim_{n\to\infty}} A_n\} = 0$, 得到矛盾. 因 $P(A_n) \geqslant 0$, 故只可能是$\sum_{n=1}^{\infty} P(A_n) = \infty$, 引理证毕.

现在讨论随机变量序列的概率 1 收敛性.

若 $\xi_n(\omega)(n=1, 2, \cdots)$, $\xi(\omega)$是随机变量, 则

$$\{\omega: \lim_{n\to\infty} \xi_n(\omega) = \xi(\omega)\}$$

$$= \left\{\omega: \bigcap_{m=1}^{\infty} \bigcup_{k=1}^{\infty} \bigcap_{n=k}^{\infty} \left(|\xi_n(\omega) - \xi(\omega)| < \frac{1}{m}\right)\right\} \tag{11}$$

这个式子可以这样理解：因为 $\omega\in\{\lim\limits_{n\to\infty}\xi_n(\omega)=\xi(\omega)\}$ 的充要条件对任一正整数 m，存在一个正整数 N，使当 $n>N$ 时均有 $|\xi_n(\omega)-\xi(\omega)|<\frac{1}{m}$；则对任一正整数 m，ω 属于$\left(|\xi_n(\omega)-\xi(\omega)|<\frac{1}{m}\right)$的下限事件，这正是(11)的右边．从这个表达式中还可以看出，$\{\lim\limits_{n\to\infty}\xi_n(\omega)=\xi(\omega)\}$是事件，因此

$$P\{\lim_{n\to\infty}\xi_n(\omega)=\xi(\omega)\}=1 \tag{12}$$

有明确的意义，这时称$\{\xi_n(\omega)\}$**以概率 1 收敛于** $\xi(\omega)$．记为 $\xi_n(\omega)\xrightarrow{a.s.}\xi(\omega)$．

因此下面两个式子都表达了$\{\xi_n(\omega)\}$以概率 1 收敛于 $\xi(\omega)$．

$$P\left\{\bigcap_{m=1}^{\infty}\bigcup_{k=1}^{\infty}\bigcap_{n=k}^{\infty}\left(|\xi_n(\omega)-\xi(\omega)|<\frac{1}{m}\right)\right\}=1 \tag{13}$$

$$P\left\{\bigcup_{m=1}^{\infty}\bigcap_{k=1}^{\infty}\bigcup_{n=k}^{\infty}\left(|\xi_n(\omega)-\xi(\omega)|\geqslant\frac{1}{m}\right)\right\}=0 \tag{14}$$

进一步，我们要说明，$\{\xi_n(\omega)\}$以概率 1 收敛于 $\xi(\omega)$的定义也可以表达为：对任意的 $\varepsilon>0$，成立

$$P\left\{\bigcap_{k=1}^{\infty}\bigcup_{n=k}^{\infty}(|\xi_n(\omega)-\xi(\omega)|\geqslant\varepsilon)\right\}=0 \tag{15}$$

若以 A_n 记$(|\xi_n(\omega)-\xi(\omega)|\geqslant\varepsilon)$，上式表示 $\overline{\lim\limits_{n\to\infty}}A_n$ 的概率为 0，这与我们对概率 1 收敛的理解一致．不过，对于这个结论，还是给它一个严格的证明．

事实上，由于对 $\varepsilon>0$，总有

$$\left\{\bigcap_{k=1}^{\infty}\bigcup_{n=k}^{\infty}(|\xi_n(\omega)-\xi(\omega)|\geqslant\varepsilon)\right\}$$

$$\subset\left\{\bigcup_{m=1}^{\infty}\bigcap_{k=1}^{\infty}\bigcup_{n=k}^{\infty}\left(|\xi_n(\omega)-\xi(\omega)|\geqslant\frac{1}{m}\right)\right\}$$

因此由(14)可以推得(15). 反之,利用

$$P\left\{\bigcup_{m=1}^{\infty}\bigcap_{k=1}^{\infty}\bigcup_{n=k}^{\infty}\left(|\xi_n(\omega)-\xi(\omega)|\geqslant\frac{1}{m}\right)\right\}$$

$$\leqslant\sum_{m=1}^{\infty}P\left\{\bigcap_{k=1}^{\infty}\bigcup_{n=k}^{\infty}\left(|\xi_n(\omega)-\xi(\omega)|\geqslant\frac{1}{m}\right)\right\}$$

可由(15)推出(14),这就说明了两种表达法的等价性。

利用概率的连续性可知,(15)等价于

$$\lim_{k\to\infty}P\left\{\bigcup_{n=k}^{\infty}(|\xi_n(\omega)-\xi(\omega)|\geqslant\varepsilon)\right\}=0 \tag{16}$$

根据德莫根定理又知(15)等价于

$$\lim_{k\to\infty}P\left\{\bigcap_{n=k}^{\infty}(|\xi_n(\omega)-\xi(\omega)|<\varepsilon)\right\}=1 \tag{17}$$

由于

$$\{|\xi_k(\omega)-\xi(\omega)|\geqslant\varepsilon\}\subset\left\{\bigcup_{n=k}^{\infty}(|\xi_n(\omega)-\xi(\omega)|\geqslant\varepsilon)\right\}$$

因此若(16)成立,则

$$\lim_{k\to\infty}P\{|\xi_k(\omega)-\xi(\omega)|\geqslant\varepsilon\}=0$$

这样一来,我们已证得

定理 1 $\xi_n(\omega)\xrightarrow{a.s.}\xi(\omega)\Longrightarrow\xi_n(\omega)\xrightarrow{P}\xi(\omega)$.

下例说明一般不能由依概率收敛推得概率 1 收敛, 所以概率 1 收敛是比依概率收敛更强的一种收敛性.

[例 1] 取 $\Omega=(0,1]$, $\mathscr{F}$ 为(0, 1]中波雷尔点集全体所构成的 σ-域, P 为勒贝格测度, 令

$$\eta_{ki}(\omega)=\begin{cases}1, & \omega\in\left(\frac{i-1}{k},\frac{i}{k}\right] \\ 0, & \omega\overline{\in}\left(\frac{i-1}{k},\frac{i}{k}\right]\end{cases} \qquad \begin{matrix} i=1,2,\cdots,k \\ k=1,2,\cdots \end{matrix} \tag{18}$$

定义

$$\xi_1(\omega)=\eta_{11}(\omega),\ \xi_2(\omega)=\eta_{21}(\omega),\ \xi_3(\omega)=\eta_{22}(\omega)$$
$$\xi_4(\omega)=\eta_{31}(\omega),\ \xi_5(\omega)=\eta_{32}(\omega),\ \cdots$$

一般 $\xi_n(\omega)=\eta_{ki}(\omega)$, 其中 $n=i+\frac{k(k-1)}{2}$. 这样定义的 $\{\xi_n(\omega)\}$ 是一列随机变量. 但对于任何一个 $\omega\in(0,1]$, $\xi_n(\omega)$ 必有无限个 k, i 使其取值 0, 也有无限个 k, i 使其取值 1, 因此$\{\xi_n(\omega)\}$ 不是以概率 1 收敛于 0. 但是另一方面, 对任意的 $e>0$,

$$P\{|\eta_{ki}(\omega)|\geqslant e\}\leqslant\frac{1}{k}$$

当 $n\to\infty$ 时, 由 $n=\frac{k(k-1)}{2}+i\leqslant\frac{k(k-1)}{2}+k$, 知道 $k\to\infty$, 因此

$$\lim_{n\to\infty}P\{|\xi_n(\omega)|\geqslant e\}$$
$$=\lim_{n\to\infty}P\{|\eta_{ki}(\omega)|\geqslant e\}=0 \qquad \left(n=\frac{k(k-1)}{2}+i\right)$$

所以$\{\xi_n(\omega)\}$依概率收敛于 0.

不难验证, $\{\xi_n(\omega)\}$是 r-阶收敛于 0 的, 因此例 1 也提供了 r-阶收敛推不出概率 1 收敛之例.

我们以前讨论的大数定律只要求依概率收敛, 若把收敛性要求提高为以概率 1 收敛, 则得到的大数定律称为**强大数定律**. 由定理 1 可知, 若强大数定律成立, 则通常的大数定律也一定成立, 反之不然. 有时为区别起见, 把依概率收敛意义下的大数定律称为**弱大数定律**.

第一个强大数定律是由波雷尔在 1909 年对贝努里试验场合建立的.

二、波雷尔强大数定律

定理2(波雷尔) 设 μ_n 是事件 A 在 n 次独立试验中的出现次数，在每次试验中事件 A 出现的概率均为 p，那么当 $n\to\infty$ 时，

$$P\left\{\frac{\mu_n}{n}\to p\right\}=1 \tag{19}$$

[证明] 为使(19)成立，由(15)知，只须对任意的 $\varepsilon>0$，成立

$$P\left\{\bigcap_{k=1}^{\infty}\bigcup_{n=k}^{\infty}\left(\left|\frac{\mu_n}{n}-p\right|\geqslant\varepsilon\right)\right\}=0 \tag{20}$$

若记 $A_n=\left\{\left|\frac{\mu_n}{n}-p\right|\geqslant\varepsilon\right\}$，则上式可写成 $P\left(\overline{\lim_{n\to\infty}}A_n\right)=0$. 根据波雷尔-康特立引理，为证明(20)只要能证明级数

$$\sum_{n=1}^{\infty}P\left\{\left|\frac{\mu_n}{n}-p\right|\geqslant\varepsilon\right\} \tag{21}$$

对任何 $\varepsilon>0$ 都收敛就可以了.

假如象证明贝努里大数定律那样用车贝晓夫不等式进行估计，只能得到

$$P\left\{\left|\frac{\mu_n}{n}-p\right|\geqslant\varepsilon\right\}\leqslant\frac{1}{4n\varepsilon^2} \tag{22}$$

这对证明弱大数定律足够了，但为了保证(21)收敛还不行，这时必须寻找更好的估计式. 在这种特殊场合，马尔可夫不等式就够用了. 由于

$$P\left\{\left|\frac{\mu_n}{n}-p\right|\geqslant\varepsilon\right\}\leqslant\frac{1}{\varepsilon^4}E\left|\frac{\mu_n}{n}-p\right|^4 \tag{23}$$

问题是要计算 $\frac{\mu_n}{n}$ 的四阶中心矩. 还是象过去一样，我们把 μ_n 表示成独立贝努里 0—1 变量之和，这样

$$\frac{\mu_n}{n}-p=\frac{1}{n}\sum_{i=1}^{n}(\xi_i-p)$$

所以

$$E\left(\frac{\mu_n}{n}-p\right)^4$$

$$=\frac{1}{n^4}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n}E(\xi_i-p)(\xi_j-p)(\xi_k-p)(\xi_l-p)$$

注意到各 ξ_i 的独立性及 $E(\xi_i-p)=0$，因此上面的和式中只有 $E(\xi_i-p)^4$ 及 $E(\xi_i-p)^2(\xi_j-p)^2$ 的项才不等于 0，显然

$$E(\xi_i-p)^4=pq(p^3+q^3) \tag{24}$$

$$E(\xi_i-p)^2(\xi_j-p)^2=p^2q^2 \quad (i\neq j) \tag{25}$$

(24)形式的项有 n 项，(25)形式的项有$\binom{4}{2}\binom{n}{2}=3n(n-1)$ 项，因此

$$E\left(\frac{\mu_n}{n}-p\right)^4=\frac{pq}{n^4}[n(p^3+q^3)+3pq(n^2-n)]<\frac{1}{4n^2} \tag{26}$$

于是

$$P\left\{\left|\frac{\mu_n}{n}-p\right|\geqslant\varepsilon\right\}<\frac{1}{4\varepsilon^4n^2} \tag{27}$$

这个估计式已经比(22)进了一步，它可以保证(21)收敛，从而证明了定理．

从本书第一节介绍随机事件频率稳定性时，我们就期待着这样一个结论，即当试验次数无限增加时，频率将趋于概率，波雷尔强大数定律正给出了这个结果．从贝努里大数定律并不能引伸出这个结论，它只断言一个不等式$\left|\frac{\mu_n}{n}-p\right|<\varepsilon$ 成立的概率可以大于 $1-\eta$，不论 η 是什么正数；但是事件：

$$\left|\frac{\mu_{n+1}}{n+1}-p\right|>\varepsilon,\ \left|\frac{\mu_{n+2}}{n+2}-p\right|>\varepsilon,\ \cdots,\ \left|\frac{\mu_{2n}}{2n}-p\right|>\varepsilon,\ \cdots$$

中至少有一个发生仍是可能的，因为它是可列个事件之并，而我们

只知道每个事件的概率很小．但波雷尔强大数定律则断言$\frac{\mu_n}{n}-p$以概率1变得很小，而且保持很小．虽然从逻辑上讲，在投硬币时每次都出现正面是可能的，这时$\frac{\mu_n}{n}=1$，因而$\frac{\mu_n}{n}\to p$并不成立，但是强大数定律断言了这种事件发生的概率为0.

三、柯尔莫哥洛夫强大数定律

下面讨论更一般的强大数定律，先把其含义进一步明确如下：

设$\{\xi_i\}$是独立随机变量序列，若

$$P\left\{\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}(\xi_i-E\xi_i)=0\right\}=1 \tag{28}$$

则称它满足强大数定律．

根据(16)，这等价于要求对任意$\varepsilon>0$成立

$$\lim_{m\to\infty}P\left\{\bigcup_{j=m}^{\infty}\left(\left|\frac{1}{j}\sum_{i=1}^{j}(\xi_i-E\xi_i)\right|\geqslant\varepsilon\right)\right\}=0 \tag{29}$$

由于

$$\bigcup_{j=m}^{\infty}\left(\left|\frac{1}{j}\sum_{i=1}^{j}(\xi_i-E\xi_i)\right|\geqslant\varepsilon\right)$$

$$\subset\left\{\sup_{j\geqslant m}\left|\frac{1}{j}\sum_{i=1}^{j}(\xi_i-E\xi_i)\right|\geqslant\varepsilon\right\} \tag{30}$$

因此需要对概率$P\left\{\sup_{j\geqslant m}\left|\frac{1}{j}\sum_{i=1}^{j}(\xi_i-E\xi_i)\right|\geqslant\varepsilon\right\}$进行估计，这相当于在独立和场合对车贝晓夫不等式进行推广．这方面已经有不少成果，在这里我们介绍一个由哈依克(Hājek)及瑞尼(Rēnyi)证明的不等式．

哈依克-瑞尼不等式　若$\{\xi_i\}$是独立随机变量序列，$D\xi_i=\sigma_i^2$

$<\infty$, $(i=1,2,\cdots)$, 而$\{C_n\}$是一列正的非增常数序列，则对任意正整数 $m,n(m<n)$及 $\varepsilon>0$, 均有

$$P\left\{\max_{m\leqslant j\leqslant n} C_j\left|\sum_{i=1}^{j}(\xi_i-E\xi_i)\right|\geqslant\varepsilon\right\}$$

$$\leqslant\frac{1}{\varepsilon^2}\left(C_m^2\sum_{j=1}^{m}\sigma_j^2+\sum_{j=m+1}^{n}C_j^2\sigma_j^2\right)\tag{31}$$

［证明］ 记

$$S_k=\sum_{j=1}^{k}(\xi_j-E\xi_j)$$

及

$$\eta=\sum_{k=m}^{n-1}S_k^2(C_k^2-C_{k+1}^2)+C_n^2S_n^2\tag{32}$$

因此

$$\eta=\sum_{k=m}^{n}S_k^2C_k^2-\sum_{k=m}^{n-1}S_k^2C_{k+1}^2$$

$$=S_m^2C_m^2+\sum_{k=m+1}^{n}(S_k^2-S_{k-1}^2)C_k^2$$

利用$\{\xi_i\}$的独立性

$$ES_k^2=\sum_{j=1}^{k}\sigma_j^2$$

所以有

$$E\eta=C_m^2\sum_{j=1}^{m}\sigma_j^2+\sum_{j=m+1}^{n}C_j^2\sigma_j^2\tag{33}$$

对 $j=m,m+1,\cdots,n$, 记

$$E_j=\{C_k|S_k|<\varepsilon,\ m\leqslant k<j;\ C_j|S_j|\geqslant\varepsilon\}\tag{34}$$

这样定义的 $E_j(j=m,m+1,\cdots,n)$是互不相容的，而且

$$P\left\{\max_{m\leq j\leq n} C_j\left|\sum_{i=1}^{j}(\xi_i-E\xi_i)\right|\geq\varepsilon\right\}$$

$$=P\left\{\max_{m\leq j\leq n} C_j|S_j|\geq\varepsilon\right\}=\sum_{j=m}^{n}P(E_j) \tag{35}$$

令

$$\chi_j(\omega)=\begin{cases}1, & \omega\in E_j\\ 0, & \omega\bar{\in} E_j\end{cases} \qquad j=m, m+1, \cdots, n \tag{36}$$

即 χ_j 是 E_j 的示性函数，注意到 E_j 的互不相容性及 $\sum_{j=m}^{n}E_j\subset\Omega$，因而

$$\sum_{j=m}^{n}\chi_j(\omega)\leqslant 1 \tag{37}$$

所以有

$$E\eta\geqslant\sum_{j=m}^{n}E(\eta\chi_j) \tag{38}$$

对 $j<k\leqslant n$，

$$S_k=S_j+(\xi_{j+1}-E\xi_{j+1})+\cdots+(\xi_k-E\xi_k)$$

因此

$$\begin{aligned}E(S_k^2\chi_j)&=E(S_j^2\chi_j)\\&\quad+E\{[(\xi_{j+1}-E\xi_{j+1})+\cdots+(\xi_k-E\xi_k)]^2\cdot\chi_j\}\\&\quad+2E\{S_j[(\xi_{j+1}-E\xi_{j+1})+\cdots+(\xi_k-E\xi_k)]\cdot\chi_j\}\\&\geqslant E(S_j^2\chi_j)+2E\{S_j\chi_j[(\xi_{j+1}-E\xi_{j+1})+\\&\quad\cdots+(\xi_k-E\xi_k)]\}\end{aligned} \tag{39}$$

由于 $S_j\chi_j$ 只与 $\xi_1, \cdots, \xi_j$ 有关，因此与$[(\xi_{j+1}-E\xi_{j+1})+\cdots+(\xi_k-E\xi_k)]$独立，故

$$E\{S_j\chi_j[(\xi_{j+1}-E\xi_{j+1})+\cdots+(\xi_k-E\xi_k)]\}$$

$$=E(S_j\chi_j)\cdot E[(\xi_{j+1}-E\xi_{j+1})+\cdots+(\xi_k-E\xi_k)]=0 \quad (40)$$

而在 E_j 上 $|S_j|\geqslant\frac{\varepsilon}{C_j}$，故有

$$E(S_j^2\chi_j)\geqslant\frac{\varepsilon^2}{C_j^2}E\chi_j=\frac{\varepsilon^2}{C_j^2}P(E_j) \quad (41)$$

因此由(39)，(40)及(41)得到

$$E(S_k^2\chi_j)\geqslant\frac{\varepsilon^2}{C_j^2}P(E_j),\quad j\leqslant k\leqslant n \quad (42)$$

现在当 $m\leqslant j\leqslant n$ 时，由(32)知

$$\begin{aligned}E(\eta\chi_j)&=\sum_{k=m}^{n-1}E(S_k^2\chi_j)(C_k^2-C_{k+1}^2)+C_n^2E(S_n^2\chi_j)\\&\geqslant\sum_{k=j}^{n-1}E(S_k^2\chi_j)(C_k^2-C_{k+1}^2)+C_n^2E(S_n^2\chi_j)\\&\geqslant\frac{\varepsilon^2}{C_j^2}P(E_j)\left[\sum_{k=j}^{n-1}(C_k^2-C_{k+1}^2)+C_n^2\right]\\&=\varepsilon^2P(E_j)\end{aligned} \quad (43)$$

由(38)及(43)

$$E\eta\geqslant\varepsilon^2\sum_{j=m}^{n}P(E_j) \quad (44)$$

利用(33)，(35)及(44)即得不等式(31)。

在曦依克-瑞尼不等式中，特别令 $m=1, C_j=1$，则得到著名的柯尔莫哥洛夫不等式。

柯尔莫哥洛夫不等式 设 $\xi_1,\xi_2,\cdots,\xi_n$ 是独立随机变量，方差有限，则对任意 $\varepsilon>0$，成立

$$P\left\{\max_{1\leqslant j\leqslant n}\left|\sum_{i=1}^{j}(\xi_i-E\xi_i)\right|\geqslant\varepsilon\right\}\leqslant\frac{1}{\varepsilon^2}\sum_{j=1}^{n}D\xi_j \quad (45)$$

柯尔莫哥洛夫不等式是概率论中最重要的不等式之一，有广

泛的应用．在上式中，若令 $n=1$，则得到

$$P\{|\xi_1-E\xi_1|\geqslant\varepsilon\}\leqslant\frac{D\xi_1}{\varepsilon^2}$$

这正是车贝晓夫不等式．因此柯尔莫哥洛夫不等式是车贝晓夫不等式的推广，而噶依克-瑞尼不等式又是柯尔莫哥洛夫不等式的推广．

利用噶依克-瑞尼不等式，能证明下面重要结果．

定理 3（柯尔莫哥洛夫强大数定律） 设 $\{\xi_i\}$，$i=1,2,\cdots$ 是独立随机变量序列，且 $\sum_{n=1}^{\infty}\frac{D\xi_n}{n^2}<\infty$，则成立

$$P\left\{\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}(\xi_i-E\xi_i)=0\right\}=1 \tag{46}$$

［证明］ 在噶依克-瑞尼不等式中，令 $C_j=\frac{1}{j}$，可以得到

$$P\left\{\max_{m\leqslant j\leqslant n}\left|\frac{1}{j}\sum_{i=1}^{j}(\xi_i-E\xi_i)\right|\geqslant\varepsilon\right\}$$

$$\leqslant\frac{1}{\varepsilon^2}\left(\frac{1}{m^2}\sum_{j=1}^{m}D\xi_j+\sum_{j=m+1}^{n}\frac{D\xi_j}{j^2}\right)$$

由概率的连续性

$$P\left\{\sup_{j\geqslant m}\left|\frac{1}{j}\sum_{i=1}^{j}(\xi_i-E\xi_i)\right|\geqslant\varepsilon\right\}$$

$$=\lim_{n\to\infty}P\left\{\max_{m\leqslant j\leqslant n}\left|\frac{1}{j}\sum_{i=1}^{j}(\xi_i-E\xi_i)\right|\geqslant\varepsilon\right\}$$

$$\leqslant\frac{1}{\varepsilon^2}\left(\frac{1}{m^2}\sum_{j=1}^{m}D\xi_j+\sum_{j=m+1}^{\infty}\frac{D\xi_j}{j^2}\right) \tag{47}$$

因为 $\sum_{j=1}^{\infty}\frac{D\xi_j}{j^2}<\infty$，故由(47)得到

$$\lim_{m\to\infty} P\left\{\sup_{j\geq m}\left|\frac{1}{j}\sum_{i=1}^{j}(\xi_i - E\xi_i)\right|\geq\varepsilon\right\}=0$$

这正等价于(46). 定理证毕.

柯尔莫哥洛夫强大数定律的另外两种常见的证明方法见习题 51～55.

显然，由柯尔莫哥洛夫强大数定律很容易推出波雷尔强大数定律.

四、独立同分布场合的强大数定律

在这种特殊的场合，可以找到强大数定律成立的充要条件，这个结果也属于柯尔莫哥洛夫.

定理 4（柯尔莫哥洛夫） 设 $\xi_1, \xi_2, \cdots$是相互独立相同分布的随机变量序列，则

$$\frac{1}{n}(\xi_1+\xi_2+\cdots+\xi_n)\xrightarrow{a.s.}a \tag{48}$$

成立的充要条件是 $E\xi_i$ 存在且等于 a.

［证明］ 若 ξ 的分布函数为 $F(x)$，我们来证明不等式:

$$\sum_{n=1}^{\infty}P\{|\xi|\geq n\}\leq E|\xi|\leq 1+\sum_{n=1}^{\infty}P\{|\xi|\geq n\} \tag{49}$$

事实上，

$$E|\xi|=\int_{-\infty}^{\infty}|x|dF(x)=\sum_{k=0}^{\infty}\int_{k\leq|x|<k+1}|x|dF(x)$$

因此

$$\sum_{k=0}^{\infty}kP\{k\leq|\xi|<k+1\}\leq E|\xi|$$

$$\leq\sum_{k=0}^{\infty}(k+1)P\{k\leq|\xi|<k+1\}$$

现在有

$$\sum_{k=0}^{\infty} kP\{k \leqslant |\xi| < k+1\} = \sum_{n=1}^{\infty} \sum_{k=n}^{\infty} P\{k \leqslant |\xi| < k+1\}$$

$$= \sum_{n=1}^{\infty} P\{|\xi| \geqslant n\}$$

及

$$\sum_{k=0}^{\infty} (k+1) P\{k \leqslant |\xi| < k+1\}$$

$$= \sum_{k=0}^{\infty} kP\{k \leqslant |\xi| < k+1\} + 1 = \sum_{n=1}^{\infty} P\{|\xi| \geqslant n\} + 1$$

这就证得了 (49). 这个不等式说明 $E|\xi| < \infty$ 的充要条件为 $\sum_{n=1}^{\infty} P\{|\xi| \geqslant n\} < \infty$.

记 $S_n = \xi_1 + \xi_2 + \cdots + \xi_n$, 若 $\frac{S_n}{n} \xrightarrow{a.s.} \mu$, 这里 μ 是有限数, 则

$$\frac{\xi_n}{n} = \frac{S_n}{n} - \frac{n-1}{n} \cdot \frac{S_{n-1}}{n-1} \xrightarrow{a.s.} 0 \tag{50}$$

这样一来, 事件 $\{|\xi_n| \geqslant n\}$ 发生无穷多次的概率为 0, 因此注意到 ξ_i 的独立性, 并利用波雷尔-康特立引理(ii), 可知

$$\sum_{n=1}^{\infty} P\{|\xi_n| \geqslant n\} < \infty$$

再由(49)即知 $E|\xi_i| < \infty$, 这时显然有 $a = E\xi_i$, 这样, 我们已证得必要性.

下证充分性. 用"截尾法", 令

$$\xi_n^* = \begin{cases} \xi_n, & |\xi_n| < n \\ 0, & |\xi_n| \geqslant n \end{cases} \tag{51}$$

先验证 $\{\xi_n^*\}$ 满足柯尔莫哥洛夫强大数定律条件. 以 $F(x)$ 记 ξ_i 的

分布函数，则

$$D\xi_n^* \leqslant E\xi_n^{*2} = \int_{-n}^{n} x^2 dF(x) \leqslant \sum_{k=1}^{n} k^2 P\{k-1 \leqslant |\xi_n| < k\}$$

$$\sum_{n=1}^{\infty} \frac{D\xi_n^*}{n^2} \leqslant \sum_{n=1}^{\infty} \sum_{k=1}^{n} \frac{k^2}{n^2} P\{k-1 \leqslant |\xi_n| < k\}$$

$$= \sum_{k=1}^{\infty} \sum_{n=k}^{\infty} \frac{k^2}{n^2} P\{k-1 \leqslant |\xi_n| < k\}$$

$$= \sum_{k=1}^{\infty} k^2 P\{k-1 \leqslant |\xi_n| < k\} \sum_{n=k}^{\infty} \frac{1}{n^2}$$

由于

$$\sum_{n=k}^{\infty} \frac{1}{n^2} < \frac{1}{k^2} + \sum_{n=k+1}^{\infty} \frac{1}{n(n-1)} = \frac{1}{k^2} + \frac{1}{k} \leqslant \frac{2}{k}$$

故

$$\sum_{n=1}^{\infty} \frac{D\xi_n^*}{n^2} < 2 \sum_{k=1}^{\infty} k P\{k-1 \leqslant |\xi_n| < k\} < \infty$$

因此

$$P\{\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} (\xi_i^* - E\xi_i^*) = 0\} = 1 \tag{52}$$

因为

$$E\xi_n^* = \int_{-n}^{n} x dF(x)$$

显然 $\lim\limits_{n\to\infty} E\xi_n^* = E\xi_1 = a$，因此 $\lim\limits_{n\to\infty} \frac{1}{n} \sum\limits_{i=1}^{n} E\xi_i^* = E\xi_1 = a$，由于

$$\left|\frac{1}{n} \sum_{i=1}^{n} (\xi_i - a)\right| \leqslant \left|\frac{1}{n} \sum_{i=1}^{n} (\xi_i - \xi_i^*)\right|$$

$$+ \left|\frac{1}{n} \sum_{i=1}^{n} (\xi_i^* - E\xi_i^*)\right| + \left|\frac{1}{n} \sum_{i=1}^{n} (E\xi_i^* - a)\right| \tag{53}$$

为证(48)成立，只须再证$\frac{1}{n}\sum_{i=1}^{n}(\xi_i-\xi_i^*)\xrightarrow{a.s.}0$. 然而

$$\sum_{i=1}^{\infty}P\{\xi_i\neq\xi_i^*\}=\sum_{i=1}^{\infty}P\{|\xi_i|\geqslant i\}\leqslant E|\xi_1|<\infty$$

由波雷尔-康特立引理知，以概率 1 有

$$\xi_i(\omega)\neq\xi_i^*(\omega), \text{只对有限个 } i \text{ 成立}$$

因此

$$P\left\{\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}(\xi_i-\xi_i^*)=0\right\}=1$$

这样，定理的证明已经完成.

显然，柯尔莫哥洛夫的这个结果是辛钦大数定律的加强，只有它才能保证在每次试验中当$n\to\infty$时，子样的均值$\frac{1}{n}\sum_{i=1}^{n}\xi_i$将最终地趋于母体的均值；当然从逻辑上讲也有可能失败，但是这种不愉快场合发生的概率等于 0. 用蒙特卡洛方法计算积分所需要的正是强大数定律.

*§5. 中心极限定理

一、林德贝格条件与费勒(Feller)条件

本节将最后解决古典的中心极限定理. 为此先把问题的提法作进一步明确.

古典的中心极限定理讨论的是独立和的分布函数向正态分布收敛的最普遍条件. 这个问题一方面可以看作是德莫哇佛-拉普拉斯古典结果的一般化，另一方面也解释了正态分布为什么是最常见的一种分布.

自从高斯指出测量误差服从正态分布之后，人们发现，正态分

布在自然界中极为常见. 例如炮弹的弹落点服从正态分布, 人的许多生理特征如身长、体重等也服从正态分布. 观察表明, 如果一个量是由大量相互独立的随机因素的影响所造成, 而每一个别因素在总影响中所起的作用不很大, 则这种量通常都服从或近似服从正态分布.

另外, 在数理统计中, 经常都假定母体服从正态分布, 这也要求通过对中心极限定理的研究来阐明假定的正确性和适用条件.

现在, 这个问题从某种意义上来讲已经得到了最后解决. 1922年林德贝格提出了充分条件; 1935年, 费勒进一步指出, 在某种条件下, 这个条件也是必要的. 这样就搞清了向正态分布收敛的充要条件. 下面就介绍这些条件.

设 $\xi_1, \xi_2, \cdots, \xi_n, \cdots$ 是一个相互独立的随机变量序列, 它们具有有限的数学期望和方差:

$$a_k = E\xi_k, \quad b_k^2 = D\xi_k \qquad (k=1, 2, \cdots, n, \cdots)$$

记

$$B_n^2 = \sum_{k=1}^{n} b_k^2$$

作和数

$$\zeta_n = \sum_{k=1}^{n} \frac{\xi_k - a_k}{B_n} \tag{1}$$

我们需要寻找和数 ζ_n 的分布函数趋于正态分布函数的充要条件.

与独立同分布场合比较, 这里保留了独立性的假定, 但是去掉了同分布的要求. 今后我们将以 $F_k(x)$ 记 ξ_k 的分布函数. 显然为了讨论 ζ_n 的极限分布, 要使问题的提法有意义, 对各个加项必须有一定要求. 例如若允许从第二项起都等于 0, 则极限分布显然由 $F_1(x)$ 完全确定, 这时就很难有什么有意思的结果. 排除这个困难的办法是规定加项中不能有某些项起支配作用, 在实际工

作中人们就是这样处理的，例如为了讨论测量的随机误差，总预先把一些系统性的误差先扣除掉.

为了使极限分布是正态分布，还要求各个加项“均匀地小”，怎样明确表达这个要求？下面先作一个启发性的推导.

设 A_k 表示下述事件:

$$\{|\xi_k-a_k|>\tau B_n\},\quad (k=1,2,\cdots,n)$$

则有

$$P\{\max_{1\leqslant k\leqslant n}|\xi_k-a_k|>\tau B_n\}=P\left\{\bigcup_{k=1}^{n}(|\xi_k-a_k|>\tau B_n)\right\}$$

$$=P\{A_1\cup A_2\cup\cdots\cup A_n\}\leqslant\sum_{k=1}^{n}P(A_k)=\sum_{k=1}^{n}\int_{|x-a_k|>\tau B_n}dF_k(x)$$

$$\leqslant\frac{1}{(\tau B_n)^2}\sum_{k=1}^{n}\int_{|x-a_k|>\tau B_n}(x-a_k)^2dF_k(x)$$

$$=\frac{1}{\tau^2B_n^2}\sum_{k=1}^{n}\int_{|x-a_k|>\tau B_n}(x-a_k)^2dF_k(x)$$

因此，只要对于任何 $\tau>0$，成立

$$\lim_{n\to\infty}\frac{1}{B_n^2}\sum_{k=1}^{n}\int_{|x-a_k|>\tau B_n}(x-a_k)^2dF_k(x)=0\qquad(2)$$

就可以保证总和(1)中各加项“均匀地小”. 上述条件(2)称为**林德贝格条件**. 林德贝格证明了条件(2)是和数(1)的分布函数趋于正态分布函数的充分条件.

但是林德贝格条件不是中心极限定理成立的必要条件. 不过，费勒进一步指出，假如下面条件得到满足:

$$\lim_{n\to\infty}\max_{k\leqslant n}\frac{b_k}{B_n}=0\qquad(3)$$

则林德贝格条件也是中心极限定理成立的必要条件.

条件(3)称为费勒条件．下面考察一下费勒条件的含义．

定理1 费勒条件(3)等价于

$$\lim_{n\to\infty} B_n=\infty \tag{4}$$

$$\lim_{n\to\infty}\frac{b_n}{B_n}=0 \tag{5}$$

[证明] 若(3)成立，则由$\frac{b_n}{B_n}\leqslant\max\limits_{k\leqslant n}\frac{b_k}{B_n}$立刻得到(5)；又若$B_n\to B$，$(B<\infty)$，则因$\max\limits_{k\leqslant n}\frac{b_k}{B_n}\geqslant\frac{b_1}{B_n}$，故$\lim\limits_{n\to\infty}\max\limits_{k\leqslant n}\frac{b_k}{B_n}\geqslant\frac{b_1}{B}>0$，这与(3)矛盾，因此应有(4)．

反之，设(4)(5)成立．对任意$\varepsilon>0$，存在正整数M，使$\frac{b_k}{B_k}<\varepsilon$对一切$k>M$成立．固定$M$之后，由于(4)，可以选一个正整数$N\geqslant M$，使$\max\limits_{k\leqslant M}\frac{b_k}{B_N}<\varepsilon$，下证对一切$n\geqslant N$均有

$$\max_{k\leqslant n}\frac{b_k}{B_n}<\varepsilon \tag{6}$$

事实上，利用B_n的单调不减性，对一切$n\geqslant N\geqslant M$有

$$\max_{k\leqslant M}\frac{b_k}{B_n}\leqslant\max_{k\leqslant M}\frac{b_k}{B_N}<\varepsilon$$

$$\max_{M<k\leqslant n}\frac{b_k}{B_n}\leqslant\max_{M<k\leqslant n}\frac{b_k}{B_k}<\varepsilon$$

因此(6)成立，这就证得了(3)，定理证毕．

量$\frac{b_n}{B_n}$可以看作是分量ξ_n对总和ζ_n的贡献，因此费勒条件相当于说：总和是大量"可忽略的"分量之和．

下面我们转入证明主要定理．

二、林德贝格-费勒定理

为了不打断主要定理的证明，我们把在定理证明中要用到的

若干事实，以引理的形式给出。

引理 1 对 $n=1,2,\cdots$ 及任意的 t,

$$\left|e^{it}-1-\frac{it}{1!}-\cdots-\frac{(it)^{n-1}}{(n-1)!}\right|\leqslant\frac{|t|^n}{n!} \tag{7}$$

［证明］记

$$g_n(t)=e^{it}-1-\frac{it}{1!}-\cdots-\frac{(it)^{n-1}}{(n-1)!}$$

先设 $t>0$，由于

$$g_1(t)=i\int_0^t e^{ix}dx \tag{8}$$

因此 $|g_1(t)|\leqslant t$，其次，对 $n>1$

$$g_n(t)=i\int_0^t g_{n-1}(x)dx \tag{9}$$

用归纳法即得(7)。

由于 $|\overline{g_n(t)}|=|g_n(t)|$，因此(7)对 $t<0$ 也成立。

特别地，我们要用到

$$|e^{it}-1-it|\leqslant\frac{|t|^2}{2} \tag{10}$$

$$\left|e^{it}-1-it+\frac{t^2}{2}\right|\leqslant\frac{|t|^3}{6} \tag{11}$$

类似地可以得到，对 $t>0$

$$1-\cos t=\left|\int_0^t \sin x\,dx\right|\leqslant\int_0^t|\sin x|dx\leqslant\int_0^t xdx=\frac{t^2}{2} \tag{12}$$

这式两边都是 t 的偶函数，故显然对 $t\leqslant0$ 也成立。

引理 2 对于任何满足 $|a_k|\leqslant1$ 及 $|b_k|\leqslant1$ $(k=1,2,\cdots,n)$ 的复数，有

$$|a_1a_2\cdots a_n-b_1b_2\cdots b_n|\leqslant\sum_{k=1}^n|a_k-b_k| \tag{13}$$

［证明］ 显然

$$a_1a_2-b_1b_2=(a_1-b_1)a_2+(a_2-b_2)b_1$$

因此

$$|a_1a_2-b_1b_2|\leqslant|a_1-b_1|+|a_2-b_2|$$

用归纳法即得(13)。

引理 3 若 $\varphi(t)$是特征函数，则 $e^{\varphi(t)-1}$ 也是特征函数，特别地

$$|e^{\varphi(t)-1}|\leqslant 1 \tag{14}$$

［证明］ 定义随机变量

$$\eta=\xi_1+\xi_2+\cdots+\xi_\nu$$

其中 $\xi_1,\xi_2,\cdots$相互独立，均有特征函数 $\varphi(t)$，ν 服从参数 $\lambda=1$ 的普阿松分布，且与诸 ξ_i 独立，不难验证 η 的特征函数为 $e^{\varphi(t)-1}$，由特征函数的性质即知(14)成立。

现在叙述并证明主要的结果。

定理 2 对(1)中定义的和数 ζ_n，成立

$$\lim_{n\to\infty}P\{\zeta_n<x\}=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-t^2/2}dt \tag{15}$$

与费勒条件(3)的充要条件是林德贝格条件(2)成立。

［证明］ 为书写方便起见，我们引用记号

$$\xi_{nk}=\frac{\xi_k-a_k}{B_n} \tag{16}$$

显然

$$E\xi_{nk}=0,\qquad D\xi_{nk}=\frac{D\xi_k}{B_n^2}=\frac{b_k^2}{B_n^2} \tag{17}$$

$$\sum_{k=1}^{n}D\xi_{nk}=\frac{1}{B_n^2}\sum_{k=1}^{n}D\xi_k=1 \tag{18}$$

以 $f_{nk}(t)$及 $F_{nk}(x)$分别表示 ξ_{nk} 的特征函数与分布函数，那么

$$F_{nk}(x)=P\left\{\frac{\xi_k-a_k}{B_n}<x\right\}=F_k(B_nx+a_k) \tag{19}$$

这时

$$\frac{1}{B_n^2}\int_{|x-a_k|>\tau B_n}(x-a_k)^2 dF_k(x)$$

$$=\int_{\left|\frac{x-a_k}{B_n}\right|>\tau}\left(\frac{x-a_k}{B_n}\right)^2 dF_k(x)=\int_{|y|>\tau}y^2 dF_{nk}(y)$$

因此林德贝格条件(2)化为:对任意 $\tau>0$,

$$\lim_{n\to\infty}\sum_{k=1}^{n}\int_{|x|>\tau}x^2 dF_{nk}(x)=0 \tag{20}$$

现在开始证明定理. 设 t 是任意固定的实数.

为证(15)必须证明

$$f_{n1}(t)\cdots f_{nn}(t)\longrightarrow e^{-t^2/2} \tag{21}$$

我们先证明,在费勒条件(3)成立的假定下,(21)与下式是等价的.

$$\sum_{k=1}^{n}(f_{nk}(t)-1)+\frac{1}{2}t^2\longrightarrow 0 \tag{22}$$

事实上,由(17)知 $f'_{nk}(0)=0$,又因为

$$f''_{nk}(t)=\int_{-\infty}^{\infty}(ix)^2 e^{itx}dF_{nk}(x)$$

故对一切 t

$$|f''_{nk}(t)|\leqslant\int_{-\infty}^{\infty}x^2 dF_{nk}(x)=\frac{b_k^2}{B_n^2}$$

把 $f_{nk}(t)$ 在原点附近展开,得到

$$f_{nk}(t)=f_{nk}(0)+f'_{nk}(0)t+\frac{1}{2}f''_{nk}(\theta t)t^2 \quad (0\leqslant\theta\leqslant 1)$$

因若(3)成立,则对任意 $\varepsilon>0$,只要 n 充分大,均有

$$\frac{b_k}{B_n}<\varepsilon,\quad k=1,2,\cdots,n \tag{23}$$

这时

$$|f_{nk}(t)-1|=\frac{1}{2}|f''_{nk}(\theta t)|t^2$$

$$\leqslant\frac{1}{2}\frac{b_k^2}{B_n^2}t^2<\frac{1}{2}\varepsilon^2 t^2 \tag{24}$$

对任意 $\delta>0$, 只要 $|z|$ 充分小, 就可以有

$$|e^z-1-z|<\delta|z| \tag{25}$$

因此由引理 3、引理 2 及(24)、(25), 只要 n 充分大, 就有

$$\left|e^{\sum_{k=1}^{n}[f_{nk}(t)-1]}-f_{n1}(t)\cdots f_{nn}(t)\right|$$

$$\leqslant\sum_{k=1}^{n}|e^{f_{nk}(t)-1}-f_{nk}(t)|$$

$$\leqslant\delta\sum_{k=1}^{n}|f_{nk}(t)-1|$$

$$\leqslant\frac{1}{2}\delta t^2\sum_{k=1}^{n}\frac{b_k^2}{B_n^2}=\frac{1}{2}\delta t^2 \tag{26}$$

因为 δ 可以任意小, 故左边趋于 0, 因此证得(21)与(22)的等价性.

接着证明定理的充分性. 先证由林德贝格条件可以推出费勒条件. 事实上,

$$\frac{b_k^2}{B_n^2}=\int_{-\infty}^{\infty}x^2dF_{nk}(x)=\int_{|x|\leqslant\tau}x^2dF_{nk}(x)+\int_{|x|>\tau}x^2dF_{nk}(x)$$

$$\leqslant\tau^2+\sum_{k=1}^{n}\int_{|x|>\tau}x^2dF_{nk}(x) \tag{27}$$

右边与 k 无关，而且 τ 可选得任意地小；对选定的 τ，由林德贝格条件(20)知道第二式当 n 足够大时也可任意地小．这样，费勒条件成立．

其次证明林德贝格条件能保证(15)成立．注意到(17)及(18)，可知

$$\sum_{k=1}^{n}[f_{nk}(t)-1]+\frac{1}{2}t^2$$

$$=\sum_{k=1}^{n}\int_{-\infty}^{\infty}\left[e^{itx}-1-itx+\frac{t^2x^2}{2}\right]dF_{nk}(x)$$

利用(11)，当 $|x|\leqslant\tau$ 时，

$$\left|e^{itx}-1-itx+\frac{t^2x^2}{2}\right|\leqslant\frac{|tx|^3}{6}\leqslant\frac{\tau|t|^3x^2}{6}$$

又利用(10)，当 $|x|>\tau$ 时，

$$\left|e^{itx}-1-itx+\frac{t^2x^2}{2}\right|\leqslant|e^{itx}-1-itx|+\frac{t^2x^2}{2}\leqslant t^2x^2$$

因此

$$\left|\sum_{k=1}^{n}[f_{nk}(t)-1]+\frac{1}{2}t^2\right|$$

$$\leqslant\sum_{k=1}^{n}\int_{|x|\leqslant\tau}\frac{\tau|t|^3x^2}{6}dF_{nk}(x)+\sum_{k=1}^{n}\int_{|x|>\tau}t^2x^2dF_{nk}(x)$$

$$\leqslant\frac{\tau|t|^3}{6}\sum_{k=1}^{n}\int_{-\infty}^{\infty}x^2dF_{nk}(x)+t^2\sum_{k=1}^{n}\int_{|x|>\tau}x^2dF_{nk}(x)$$

$$=\frac{\tau|t|^3}{6}+t^2\sum_{k=1}^{n}\int_{|x|>\tau}x^2dF_{nk}(x)\qquad(28)$$

对任给的 $\varepsilon>0$，由于 τ 的任意性，可选得使 $\frac{\tau|t|^3}{6}<\frac{\varepsilon}{2}$，对选定的 τ，用林德贝格条件知只要 n 充分大，也可使 $t^2\sum_{k=1}^{n}\int_{|x|>\tau}x^2dF_{nk}(x)<\frac{\varepsilon}{2}$，因此我们已证得了(22)，但由于我们已证过费勒条件(3)成立，这时(22)与(21)是等价的，因而(21)也成立，根据特征函数连续性定理可知(15)成立。

再证定理的必要性。

由于(15)成立，因此相应的特征函数应满足(21)，但在费勒条件成立时，这又推出(22)，因此

$$
\begin{aligned}
&\sum_{k=1}^{n}[f_{nk}(t)-1]+\frac{t^2}{2}\\
&=\sum_{k=1}^{n}\int_{-\infty}^{\infty}\left[e^{itx}-1+\frac{t^2x^2}{2}\right]dF_{nk}(x)\longrightarrow 0
\end{aligned}
\tag{29}
$$

因为由(12)可得 $\cos tx-1+\frac{t^2x^2}{2}\geqslant 0$，因此上述被积函数的实部是非负的，故

$$
\begin{aligned}
&\operatorname{Re}\left(\sum_{k=1}^{n}\int_{-\infty}^{\infty}\left[e^{itx}-1+\frac{t^2x^2}{2}\right]dF_{nk}(x)\right)\\
&=\sum_{k=1}^{n}\int_{-\infty}^{\infty}\left[\cos tx-1+\frac{t^2x^2}{2}\right]dF_{nk}(x)\\
&\geqslant\sum_{k=1}^{n}\int_{|x|>\tau}\left[\cos tx-1+\frac{t^2x^2}{2}\right]dF_{nk}(x)\\
&\geqslant\sum_{k=1}^{n}\int_{|x|>\tau}\left(\frac{t^2x^2}{2}-2\right)dF_{nk}(x)
\end{aligned}
$$

$$=\frac{t^2}{2}\sum_{k=1}^{n}\int_{|x|>\tau}x^2dF_{nk}(x)-2\sum_{k=1}^{n}\int_{|x|>\tau}dF_{nk}(x)$$

$$\geqslant\left(\frac{t^2}{2}-2\tau^{-2}\right)\sum_{k=1}^{n}\int_{|x|>\tau}x^2dF_{nk}(x) \tag{30}$$

因为对任意 $\tau>0$，可找到 t，使$\frac{t^2}{2}-2\tau^{-2}>0$，这时由（29），（30）可得

$$\left(\frac{t^2}{2}-2\tau^{-2}\right)\sum_{k=1}^{n}\int_{|x|>\tau}x^2dF_{nk}(x)\longrightarrow 0$$

故林德贝格条件成立，定理证毕。

三、若干推论

林德贝格条件给出了中心极限定理成立的普遍条件，由它可以推出许多特殊的结果。

首先，我们来说明独立同分布场合的林德贝格-勒维定理是定理 2 的特例。

若 $\xi_1,\xi_2,\cdots$是独立同分布随机变量序列，$E\xi_k=a$，$0<\sigma^2=D\xi_k<\infty$，则

$$B_n=\sqrt{n}\,\sigma \tag{31}$$

这时

$$\frac{1}{B_n^2}\sum_{k=1}^{n}\int_{|x-a_k|>\tau B_n}(x-a_k)^2dF_k(x)$$

$$=\frac{1}{n\sigma^2}\cdot n\int_{|x-a|>\tau\sigma\sqrt{n}}(x-a)^2dF(x) \tag{32}$$

由于方差 $0<\sigma^2<\infty$，上式右边的积分当 $n\to\infty$ 时趋于 0，故林德贝格条件得到满足，所以中心极限定理成立。

下面我们再来给出两个有用的结果。

定理 3 若$\xi_1, \xi_2, \cdots$是独立随机变量序列，存在常数K_n，使$\max\limits_{1\leqslant j\leqslant n}|\xi_j|\leqslant K_n (n=1, 2, \cdots)$，且$\lim\limits_{n\to\infty}\frac{K_n}{B_n}=0$，则

$$P\left\{\sum_{k=1}^{n}\frac{\xi_k-a_k}{B_n}<x\right\}\to\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-t^2/2}dt$$

[证明] 由假定，对任意的$\varepsilon>0$，只要n充分大就有$K_n\leqslant\varepsilon B_n$，因此

$$\{|\xi_j|\leqslant\varepsilon B_n\}=\Omega \quad (1\leqslant j\leqslant n) \tag{33}$$

故

$$\lim_{n\to\infty}\frac{1}{B_n^2}\sum_{j=1}^{n}\int_{|x_j|\leqslant\varepsilon B_n}(x-a_j)^2dF_j(x)$$

$$\lim_{n\to\infty}\frac{1}{B_n^2}\sum_{j=1}^{n}\int_{-\infty}^{\infty}(x-a_j)^2dF_j(x)=1$$

因此林德贝格条件得到满足，所以中心极限定理成立。

定理 4（李雅普洛夫） 如果对相互独立的随机变量序列$\xi_1, \xi_2, \cdots, \xi_n, \cdots$，能选择这样一个正数$\delta>0$，使当$n\to\infty$时，

$$\frac{1}{B_n^{2+\delta}}\sum_{k=1}^{n}E|\xi_k-a_k|^{2+\delta}\longrightarrow 0 \tag{34}$$

则

$$P\left\{\frac{1}{B_n}\sum_{k=1}^{n}(\xi_k-a_k)<x\right\}\to\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-t^2/2}dt$$

[证明] 只要验证林德贝格条件就行了。事实上，

$$\frac{1}{B_n^2}\sum_{k=1}^{n}\int_{|x-a_k|>\varepsilon B_n}(x-a_k)^2dF_k(x)$$

$$\leqslant \frac{1}{B_n^2(\tau B_n)^{\delta}} \sum_{k=1}^{n} \int_{|x-a_k|>\tau B_n} |x-a_k|^{2+\delta} dF_k(x)$$

$$\leqslant \frac{1}{\tau^{\delta}} \cdot \frac{1}{B_n^{2+\delta}} \sum_{k=1}^{n} \int_{-\infty}^{\infty} |x-a_k|^{2+\delta} dF_k(x) \to 0. \quad (n\to\infty)$$

用定理 2 就可推得所需的结论。

第五章小结

本章研究了极限定理，这是概率论基础中比较深入的结果；前几章学到的知识在这里得到了综合应用；一些重要问题在这里进一步讨论并获得解决。

在我们的课程中，为了使读者对极限定理有直观的认识，是从贝努里试验场合开始叙述的。这里所用的工具比较初等：贝努里大数定律是用矩法证明的；德莫哇佛-拉普拉斯定理则通过利用斯特灵公式进行渐近估计而得到。

接着我们处理独立同分布场合，这是贝努里试验的直接推广，也是在实际中，特别是数理统计中，最常碰到的情况。为了证明辛钦大数定律及林德贝格-勒维定理已用到特征函数，所用的方法具有普遍性及简明性，是读者比较容易理解的。应当指出，收敛性概念及特征函数这一工具是深入研究极限定理所不可缺少的，所以对这部分内容我们作了相当详细的叙述。后面用到的主要是结论，因此有关证明在初学时不妨略去。

最后，介绍了强大数定律及一般场合的中心极限定理，这是概率论中相当深刻的结果。前者的证明通过建立比车贝晓夫不等式更为锐利的不等式而实现；后者的证明则得力于特征函数这一有力的工具。到此为止，概率论中提出古典极限定理问题已获得了令人满意的解决。

习　题

1. 设 $f(x)(0\leq x<\infty)$ 是单调非降函数，且 $f(x)>0$. 对随机变量 ξ，若 $Ef(|\xi|)<\infty$，则对任意 $x>0$，$P\{|\xi|\geq x\}\leq\frac{1}{f(x)}Ef(|\xi|)$.

2. ξ 为非负随机变量，若 $Ee^{a\xi}<\infty,(a>0)$，则对任意 $x>0,P\{\xi\geq x\}\leq e^{-ax}Ee^{a\xi}$.

3. 若 $h(x)\geq 0,\xi$ 为随机变量，且 $Eh(\xi)<\infty$，则关于任何 $C>0,P\{h(\xi)\geq C\}\leq C^{-1}Eh(\xi)$.

4. $\{\xi_k\}$ 各以 $\frac{1}{2}$ 概率取值 k^s 和 $-k^s$，当 s 为何值时，大数定律可用于独立随机变量序列 $\xi_1,\cdots,\xi_n,\cdots$ 的算术平均值？

5. 若 ξ_k 的分布列为 $\begin{pmatrix}\sqrt{\ln k}, & -\sqrt{\ln k}\\ \frac{1}{2}, & \frac{1}{2}\end{pmatrix}$，试证大数定律适用于独立随机变量序列 $\{\xi_k\}$.

6. 验证概率分布如下给定的独立随机变量序列是否满足马尔可夫条件：

(1) $P\{X_k=\pm 2^k\}=\frac{1}{2}$;

(2) $P\{X_k=\pm 2^k\}=2^{-(2k+1)}$，$P\{X_k=0\}=1-2^{-2k}$;

(3) $P\{X_k=\pm k\}=\frac{1}{2}k^{-\frac{1}{2}}$，$P\{X_k=0\}=1-k^{-\frac{1}{2}}$.

7. 若 ξ_k 具有有限方差，服从同一分布，但各 k 间，ξ_k 和 ξ_{k+1} 有相关，而 $\xi_k,\xi_l,(|k-l|\geq 2)$ 是独立的，证明这时对 $\{\xi_k\}$ 大数定律成立.

8. （伯恩斯坦定理）已知随机变量序列 $\xi_1,\xi_2,\cdots$ 的方差有界：$D\xi_n\leq C$，并且当 $|i-j|\to\infty$ 时，相关系数 $r_{ij}\to 0$，证明对 $\{\xi_n\}$ 成立大数定律.

9. （格涅坚科定理）对随机变量序列 $\{\xi_i\}$，若记 $\eta_n=\frac{1}{n}(\xi_1+\cdots+\xi_n)$，$a_n=\frac{1}{n}(E\xi_1+\cdots+E\xi_n)$，则 $\{\xi_i\}$ 服从大数定律的充要条件是

$$\lim_{n\to\infty}E\left\{\frac{(\eta_n-a_n)^2}{1+(\eta_n-a_n)^2}\right\}=0$$

10. 用斯特灵公式证明：当 $n\to\infty, m\to\infty, n-m\to\infty$，而 $\frac{m}{n}\to 0$ 时，

$$\binom{2n}{n-m}\left(\frac{1}{2}\right)^{2n}\sim\frac{1}{\sqrt{\pi n}}e^{-m^2/n}$$

11. 用(1.33)计算 $b(5; 500, 0.01)$ 及 $b(40; 10000, 0.005)$ 并与精确值比较.

12. 某计算机系统有120个终端，每个终端有5%时间在使用，若各个终端使用与否是相互独立的，试求有10个或更多终端在使用的概率。

13. 求证，在 $x>0$ 时，有不等式

$$\frac{x}{1+x^2}e^{-x^2/2}\leqslant\int_x^\infty e^{-t^2/2}dt\leqslant\frac{1}{x}e^{-x^2/2}$$

14. 用德莫哇佛-拉普拉斯定理证明，在贝努里试验中，若 $0<p<1$，则不管 K 是如何大的常数，总有

$$P\{|\mu_n-np|<K\}\to 0\quad(n\to\infty)$$

15. 用车贝晓夫不等式确定当掷一均匀铜币时，需投多少次才能保证使得正面出现的频率在0.4至0.6之间的概率不小于90%，并用正态逼近计算同一问题.

16. 用车贝晓夫不等式及德莫哇佛-拉普拉斯极限定理估计下面概率：

$$P\left\{\left|\frac{\mu_n}{n}-p\right|\geqslant\varepsilon\right\}$$

并进行比较. 这里 μ_n 是 n 次贝努里试验中成功总次数，p 为每次成功的概率.

17. 现有一大批种子，其中良种占1/6，今在其中任选6000粒，试问在这些种子中良种所占的比例与1/6之差小于1%的概率是多少？

18. 种子中良种占1/6，我们有99%的把握断定在6000粒种子中良种所占的比例与1/6之差是多少？这时相应的良种粒数落在哪个范围内.

19. 蒲丰试验中掷铜币4040次，出正面2048次，试计算当重复蒲丰试验时，正面出现的频率与概率之差的偏离程度不大于蒲丰试验中所发生的偏差的概率.

20. 设分布函数列 $\{F_n(x)\}$ 弱收敛于连续的分布函数 $F(x)$，试证这收敛对 $x\in R^1$ 是一致的.

21. 设 $F_n(x)$ 为一列正态分布函数，收敛于分布函数 $F(x)$，试证 $F(x)$ 也是正态分布函数。

22. 试证若正态随机变量序列依概率收敛，则其数学期望与方差也收敛.

23. 若 X_n 为多维正态随机向量，$X_n \xrightarrow{P} X$，试证 X 为正态向量.

24. 若 X_n 的概率分布为$\begin{pmatrix} 0, & n \\ 1-\frac{1}{n}, & \frac{1}{n} \end{pmatrix}$，试证相应的分布函数收敛，但矩不收敛.

25. 随机变量序列 $\{\xi_n\}$ 具有分布函数 $\{F_n(x)\}$，且 $F_n(x) \to F(x)$，又 $\{\eta_n\}$依概率收敛于常数 $C>0$. 试证：(1)$\zeta_n = \xi_n + \eta_n$ 的分布函数收敛于 $F(x-C)$；(2)$\zeta_n = \frac{\xi_n}{\eta_n}$的分布函数收敛于 $F(Cx)$.

26. 试证：(1) $X_n \xrightarrow{P} X \Rightarrow X_n - X \xrightarrow{P} 0$；

(2) $X_n \xrightarrow{P} X$，$X_n \xrightarrow{P} Y \Rightarrow P\{X=Y\}=1$；

(3) $X_n \xrightarrow{P} X \Rightarrow X_n - X_m \xrightarrow{P} 0(n, m \to \infty)$；

(4) $X_n \xrightarrow{P} X$，$Y_n \xrightarrow{P} Y \Rightarrow X_n \pm Y_n \xrightarrow{P} X \pm Y$；

(5) $X_n \xrightarrow{P} X$，k 是常数$\Rightarrow kX_n \xrightarrow{P} kX$；

(6) $X_n \xrightarrow{P} X \Rightarrow X_n^2 \xrightarrow{P} X^2$；

(7) $X_n \xrightarrow{P} a, Y_n \xrightarrow{P} b$, a, b 是常数$\Rightarrow X_nY_n \xrightarrow{P} ab$；

(8) $X_n \xrightarrow{P} 1 \Rightarrow X_n^{-1} \xrightarrow{P} 1$；

(9) $X_n \xrightarrow{P} a, Y_n \xrightarrow{P} b$, a, b 是常数, $b \neq 0 \Rightarrow X_nY_n^{-1} \xrightarrow{P} ab^{-1}$；

(10) $X_n \xrightarrow{P} X$, Y 是随机变量$\Rightarrow X_nY \xrightarrow{P} XY$；

(11) $X_n \xrightarrow{P} X$，$Y_n \xrightarrow{P} Y \Rightarrow X_nY_n \xrightarrow{P} XY$.

27. 设 $X_n \xrightarrow{P} X$，而 g 是 R^1 上的连续函数，试证 $g(X_n) \xrightarrow{P} g(X)$.

28. 若$\{X_n\}$是单调下降的正随机变量序列，且 $X_n \xrightarrow{P} 0$，试证 $X_n \xrightarrow{a.s.} 0$.

29. 若 $X_1, X_2, \cdots$ 是独立随机变量序列，μ 是整值随机变量，$P\{\mu=k\}=p_k$，且与$\{X_i\}$独立，求 $\eta = X_1 + X_2 + \cdots + X_\mu$ 的特征函数.

30. 若$f(t)$是非负定函数，试证：(1)$f(0)$是实的，且$f(0)\geqslant 0$；(2)$f(-t)=\overline{f(t)}$；(3)$|f(t)|\leqslant f(0)$.

*31. 证明：(1)若 $g(x)$ 是连续函数，且$g(\mp\infty)=0$，又分布函数列$\{F_n(x)\}$弱收敛于分布函数 $F(x)$，则

$$\int_{-\infty}^{\infty} g(x)dF_n(x)\to\int_{-\infty}^{\infty} g(x)dF(x)$$

(2)若分布函数 $F(x)$ 对应的特征函数为 $f(t)$，定义**积分特征函数**为

$$\hat{f}(t)=\int_0^t f(u)du=\int_{-\infty}^{\infty}\frac{e^{itx}-1}{ix}dF(x)$$

试证在 $F(x)$ 与 $\hat{f}(t)$之间有一个一一对应；(3)若分布函数列 $\{F_n(x)\}$ 弱收敛于分布函数列 $F(x)$，则相应的积分特征函数收敛：$\hat{f}_n(t)\to\hat{f}(t)$；反之，若$\hat{f}_n(t)$ 收敛到一个函数 $\hat{h}(t)$，则存在一个分布函数使 $F_n(x)\xrightarrow{W}F(x)$，且 $\hat{f}(t)=\hat{h}(t)$.

32. 用特征函数法直接证明德莫哇佛-拉普拉斯积分极限定理.

33. 若母体 ξ 的数学期望 $E\xi=m$，$D\xi=\sigma^2$，抽容量为 n 的子样求其平均值 $\bar{\xi}$，为使 $P\{|\bar{\xi}-m|<0.1\sigma\}\geqslant 95\%$，问 n 应取多大值？

34. 若$\{\xi_n, n=1,2,\cdots\}$为相互独立随机变量序列，具有相同分布

$$P\{\xi_n=1\}=\frac{1}{2},\ P\{\xi_n=0\}=\frac{1}{2}$$

而 $\eta_n=\sum_{k=1}^{n}\frac{\xi_k}{2^k}$，试证 η_n 的分布收敛于[0,1]上的均匀分布.

35. 用特征函数法证明二项分布的普阿松逼近定理.

36. 用特征函数法证明，普阿松分布当 $\lambda\to\infty$ 时，渐近正态分布.

37. 若 X_i 是独立同分布随机变量序列，其分布分别为：(1)$[-a,a]$上均匀分布；(2)普阿松分布；(3)Γ-分布，记

$$Y_n=\frac{\sum_{i=1}^{n}(X_i-EX_i)}{\sqrt{\sum_{i=1}^{n}DX_i}}$$

试计算 Y_n 的特征函数，并求 $n\to\infty$ 时的极限.

38. 设 X_n 独立同分布，$P\{X_n=2^{k-2\log k}\}=2^{-k}(k=1, 2, \cdots)$，则大数定律成立。

39. 若$\{X_i\}$是相互独立的随机变量序列，均服从 $N(0,1)$，试证

$$W_n=\sqrt{n}\frac{X_1+\cdots+X_n}{X_1^2+\cdots+X_n^2} \quad 及 \quad U_n=\frac{X_1+\cdots+X_n}{\sqrt{X_1^2+\cdots+X_n^2}}$$

渐近正态分布$N(0,1)$。

40. 设 $X_1, X_2, \cdots$是独立随机变量序列，均服从$[0,1]$均匀分布，令

$$Z_n=\left(\prod_{i=1}^{n}X_i\right)^{1/n}$$

试证$Z_n \xrightarrow{P} C$，这里C是常数，并求C。

41. 若$\{X_i\}$是独立同分布随机变量序列，$EX_i=m$，若$f(x)$是一个有界的连续函数，试证

$$\lim_{n\to\infty}E\left[f\left(\frac{X_1+\cdots+X_n}{n}\right)\right]=f(m)$$

42. 若$\{X_i\}$是独立同分布、具有有限二阶矩的随机变量序列，试证

$$\frac{2}{n(n+1)}\cdot\sum_{i=1}^{n}iX_i \xrightarrow{P} EX_1$$

43. 设 $X_1, X_2, \cdots$相互独立，均服从柯西分布 $p(x)=\frac{1}{\pi}\cdot\frac{1}{1+x^2}$，试证它们不满足格涅坚科关于大数定律的充要条件（见本章习题 9），即要指出，当 $n\to\infty$ 时

$$E\frac{\left(\sum_{i=1}^{n}X_i\right)^2}{n^2+\left(\sum_{i=1}^{n}X_i\right)^2} \not\to 0$$

*44. （魏尔斯特拉斯定理的概率论证明） 设 $f(x)$是$[0, 1]$上连续函数，利用概率论方法证明：必存在多项式序列$\{B_n(x)\}$，在$[0, 1]$上一致收敛于 $f(x)$。（提示：定义伯恩斯坦多项式

$$B_n(x)=\sum_{m=0}^{n}\binom{n}{m}x^m(1-x)^{n-m}f\left(\frac{m}{n}\right)$$

并利用大数定律)。

45. 设$\{X_i\}$是独立随机变量序列. 试证$X_n\xrightarrow{a.s.}0$的充要条件为对任意$\varepsilon>0$, 有$\sum_{n=1}^{\infty}P\{|X_n|\geqslant\varepsilon\}<\infty$.

46. 试证独立同分布随机变量序列, 若存在有限的四阶中心矩, 则强大数定律成立.

47. 习题6的独立随机变量序列是否满足强大数定律?

48. 举例说明波雷尔-康特立引理(i)之逆不成立.

49. 设$\{X_n\}$是相互独立且具有有限方差的随机变量序列, 若

$$\sum_{n=1}^{\infty}\frac{DX_n}{n^2}<\infty$$

则必有

$$\lim_{n\to\infty}\frac{1}{n^2}\sum_{k=1}^{n}DX_k=0$$

50. 设$f(x)$和$g(x)$在闭区间$[0,1]$上连续, 且满足$0\leqslant f(x)\leqslant Cg(x)$, 这里$C$是一个正常数, 则成立

$$\lim_{n\to\infty}\int_0^1\int_0^1\cdots\int_0^1\frac{f(x_1)+f(x_2)+\cdots+f(x_n)}{g(x_1)+g(x_2)+\cdots+g(x_n)}dx_1dx_2\cdots dx_n$$
$$=\frac{\int_0^1 f(x)dx}{\int_0^1 g(x)dx}$$

*51. 直接证明柯尔莫哥洛夫不等式.

*52. 对事件序列$\{A_n\}$, 及整数序列$1=n_1<n_2<\cdots$, 证明:

(1) $\bigcap_{k=1}^{\infty}\bigcup_{n=k}^{\infty}A_n=\bigcap_{k=1}^{\infty}\bigcup_{j=k}^{\infty}\bigcup_{n=n_j}^{n_{j+1}-1}A_n$;

(2) 若$\sum_{j=1}^{\infty}P\left\{\bigcup_{n=n_j}^{n_{j+1}-1}A_n\right\}<\infty$, 则$P\{\overline{\lim_{n\to\infty}}A_n\}=0$;

(3) 对随机变量序列$\{\zeta_n\}$，若级数

$$\sum_{j=1}^{\infty} P\{\max_{n_j\leqslant n<n_{j+1}} |\zeta_n|\geqslant\varepsilon\}$$

对一切$\varepsilon>0$收敛，则$P\{\zeta_n\to 0\}=1$.

*53. 若$\{\xi_k\}$是独立随机变量序列，方差有限，记

$$S_n=\sum_{k=1}^{n}(\xi_k-E\xi_k),\quad \eta_n=\frac{1}{n}S_n$$

(1) 利用柯尔莫哥洛夫不等式证明：

$$p_m\equiv P\{\max|\eta_n|\geqslant\varepsilon, 2^m\leqslant n<2^{m+1}\}$$

$$\leqslant\frac{1}{(2^m\varepsilon)^2}\sum_{j<2^{m+1}} D\xi_j$$

(2) 对上述p_m，证明若$\sum_{k=1}^{\infty}\frac{D\xi_k}{k^2}<\infty$，则$\sum_{m=1}^{\infty}p_m$收敛；

(3) 利用上题结果证明对$\{\xi_n\}$成立柯尔莫哥洛夫强大数定律.

*54. (1) 设$\{c_k\}$为常数列，令

$$s_n=\sum_{k=1}^{n}c_k,\quad b_m=\sup\{|s_{m+k}-s_m|, k=1,2,\cdots\}$$

$$b=\inf\{b_m, m=1,2,\cdots\}$$

试证$\sum_{k=1}^{\infty}c_k$收敛的充要条件是$b=0$；(2) (Kronecker 引理)对实数列$\{c_k\}$，若$\sum_{k=1}^{\infty}\frac{c_k}{k}$收敛，则$\frac{1}{n}\sum_{k=1}^{n}c_k\to 0$.

*55. 若$\{\xi_k\}$是独立随机变量序列，方差有限，记

$$S_n'=\sum_{k=1}^{n}\frac{\xi_k-E\xi_k}{k}$$

(1) 利用柯尔莫哥洛夫不等式证明

$$P\{\max_{1\leqslant k\leqslant n}|S_{m+k}'-S_m'|\geqslant\varepsilon\}\leqslant\frac{1}{\varepsilon^2}\sum_{k=m+1}^{m+n}\frac{D\xi_k}{k^2}$$

(2) 利用上题结论(1)证明：若$\sum_{k=1}^{\infty}\frac{D\xi_k}{k^2}<\infty$，则 S_n' 以概率 1 收敛；

(3) 利用 Kronecker 引理导出柯尔莫哥洛夫强大数定律.

56. 设 $X_1, X_2, \cdots$ 是独立随机变量序列，对它成立中心极限定理，则对$\{X_n\}$成立大数定律的充要条件为 $D(X_1+\cdots+X_n)=o(n^2)$.

57. 设 $X_1, X_2, \cdots$ 是独立同分布随机变量序列，且 $\frac{\sum_{k=1}^{n}X_k}{\sqrt{n}}$ 对每一个 $n=1,2,\cdots$有相同分布，那么，若 $EX_i=0$, $DX_i=1$，则 X_i 必须是$N(0,1)$变量.

58. 设$\{X_k\}$是独立随机变量序列，且 X_k 服从$N(0,2^{-k})$，试证序列$\{X_k\}$：(1)成立中心极限定理；(2)不满足费勒条件；(3)不满足林德贝格条件，从而说明林德贝格条件并不是中心极限定理成立的必要条件.

59. 若$\{X_k\}$是独立随机变量序列，X_1 服从$[-1, 1]$均匀分布，对 $k=2, 3, \cdots$, X_k 服从 $N(0, 2^{k-1})$，证明对 $\{X_k\}$ 成立中心极限定理但不满足费勒条件.

60. 在普阿松试验中，第 i 次试验时事件 A 出现的概率为 p_i，不出现概率为 q_i，各次试验是独立的，以 ν_n 记前 n 次试验中事件 A 出现的次数. 试证：

(1) $\frac{\nu_n-E\nu_n}{n}\xrightarrow{P}0$；　　(2) 对$\frac{\nu_n-\sum_{i=1}^{n}p_i}{\sqrt{\sum_{i=1}^{n}p_iq_i}}$

成立中心极限定理的充要条件是$\sum_{i=1}^{\infty}p_iq_i=+\infty$.

61. 设$\{X_k\}$独立，X_k 服从$[-k, k]$均匀分布，问对$\{X_k\}$能否用中心极限定理？

62. 试问对下列独立随机变量序列，李雅普洛夫定理是否成立？

(1) $X_k:\begin{pmatrix}-\sqrt{k}, & \sqrt{k}\\ \frac{1}{2}, & \frac{1}{2}\end{pmatrix}$；(2) $X_k:\begin{pmatrix}-k^{\alpha}, & 0, & k^{\alpha}\\ \frac{1}{3}, & \frac{1}{3}, & \frac{1}{3}\end{pmatrix}$，$\alpha>0$.

63. 求证：当 $n\to\infty$ 时

$$\frac{\left(\frac{n}{2}\right)^{\frac{n}{2}}}{\Gamma\left(\frac{n}{2}\right)}\int_0^{1+t\sqrt{\frac{2}{n}}} z^{\frac{n}{2}-1}e^{-\frac{nz}{2}}dz\to\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{t}e^{-z^2/2}dz$$

64. 设 $\lambda>0$，独立随机变量序列 $\{X_k\}$，对一切 k，X_k 以概率 $\frac{1}{2}$ 分别取值 $\pm k^{\lambda}$，试证：当 $\lambda<\frac{1}{2}$ 时大数定律成立，对 $\lambda\geqslant 1$ 则不成立。

65. 求证：当 $n\to\infty$ 时，$e^{-n}\sum_{k=0}^{n}\frac{n^k}{k!}\to\frac{1}{2}$。

附 录 一

常用分布表

分布名称	概率分布与密度函数 $p(x)$	数学期望	方差	特征函数	有关章节
退化分布 $I(x-c)$	$p_c=1$ (c 为常数)	c	0	e^{ict}	三§1；四§4； 五§2，§3.
贝努里分布 (两点分布)	$p_k=\begin{cases}q, & k=0\\ p, & k=1\end{cases}$ $0<p<1,\ q=1-p$	p	pq	$pe^{it}+q$	二§3；三§1； 四§1；五§1.
二项分布 $b(k;\ n,p)$	$b(k;\ n,p)=\binom{n}{k}p^k q^{n-k}$ $k=0,1,\cdots,n.$ $0<p<1, q=1-p$	np	npq	$(pe^{it}+q)^n$	一§3；二§3， §4；三§1； 四§1，§3，§4； 五§1.
普阿松分布 $p(k;\ \lambda)$	$p(k;\ \lambda)=\frac{\lambda^k}{k!}e^{-\lambda}, k=0,1,2,\cdots$ $\lambda>0$	λ	λ	$e^{\lambda(e^{it}-1)}$	二§4；三§1； 四§1，§3，§4.
几何分布 $g(k;\ p)$	$g(k;\ p)=q^{k-1}p$ $k=1,2,\cdots$ $0<p<1,\ q=1-p$	$\frac{1}{p}$	$\frac{q}{p^2}$	$\frac{pe^{it}}{1-qe^{it}}$	一§3；二§3； 三§1；四§1， §3.

超几何分布	$p_k=\dfrac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}$, $M\leqslant N$, $n\leqslant N$, M,N,n 正整数 $k=0,1,2,\cdots,\min(M,N)$	$\dfrac{nM}{N}$	$\dfrac{nM}{N}\left(1-\dfrac{M}{N}\right)\cdot\dfrac{N-n}{N-1}$	$\sum\limits_{k=0}^{n}\dfrac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}e^{itk}$	一§3；三§1；四§1，§3.
巴斯卡分布	$p_k=\binom{k-1}{r-1}p^r q^{k-r}, k=r,r+1,\cdots$ $0<p<1, q=1-p$, r 正整数	$\dfrac{r}{p}$	$\dfrac{rq}{p^2}$	$\left(\dfrac{pe^{it}}{1-qe^{it}}\right)^r$	二§3；三§1.
正态分布（高斯分布）$N(a,\sigma^2)$	$p(x)=\dfrac{1}{\sqrt{2\pi}\sigma}e^{-(x-a)^2/2\sigma^2}$ $-\infty<x<\infty$, a, $\sigma>0$, 常数	a	σ^2	$e^{iat-\frac{1}{2}\sigma^2t^2}$	三§1，§2，§3；四§1，§2，§4，§5；五§1，§3，§5.
均匀分布 $U[a,b]$	$p(x)=\begin{cases}\dfrac{1}{b-a}, & a\leqslant x\leqslant b\\ 0, & 其它\end{cases}$ $a<b$, 常数	$\dfrac{a+b}{2}$	$\dfrac{(b-a)^2}{12}$	$\dfrac{e^{itb}-e^{ita}}{it(b-a)}$	一§4；三§1，§2，§3；四§1，§2；五§3.
指数分布	$p(x)=\begin{cases}\lambda e^{-\lambda x}, & x\geqslant 0\\ 0, & x<0\end{cases}$ $\lambda>0$, 常数	λ^{-1}	λ^{-2}	$\left(1-\dfrac{it}{\lambda}\right)^{-1}$	三§1，§3；四§1，§2.
χ^2-分布	$p(x)=\begin{cases}\dfrac{1}{2^{n/2}\Gamma\left(\dfrac{n}{2}\right)}x^{\frac{n}{2}-1}e^{-x/2}, & x\geqslant 0\\ 0, & x<0\end{cases}$ n 正整数	n	$2n$	$(1-2it)^{-n/2}$	三§3.

续前表

分布名称	概率分布与密度函数 $p(x)$	数学期望	方差	特征函数	有关章节
Γ-分布	$p(x)=\begin{cases}\frac{\lambda^r}{\Gamma(r)}x^{r-1}e^{-\lambda x}, & x\geq 0\\ 0, & x<0\end{cases}$ $r>0$，$\lambda>0$ 常数	$r\lambda^{-1}$	$r\lambda^{-2}$	$\left(1-\frac{it}{\lambda}\right)^{-r}$	三§1,§3.
柯西分布	$p(x)=\frac{1}{\pi}\cdot\frac{\lambda}{\lambda^2+(x-\mu)^2}$ $-\infty<x<\infty$，$\lambda>0$，μ 常数	不存在	不存在	$e^{i\mu t-\lambda\|t\|}$	三§3, 四§1.
t-分布	$p(x)=\frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)}\left(1+\frac{x^2}{n}\right)^{-(n+1)/2}$ $-\infty<x<\infty$ n 正整数	$0(n>1)$	$\frac{n}{n-2}(n>2)$		
Pareto 分布	$p(x)=\begin{cases}rA^r\frac{1}{x^{r+1}}, & x\geq A\\ 0, & x<A\end{cases}$ $r>0$，$A>0$	($r>1$ 时存在)	($r>2$ 时存在)		

F-分布	$p(x)=\begin{cases}\dfrac{\Gamma\left(\frac{k_1+k_2}{2}\right)}{\Gamma\left(\frac{k_1}{2}\right)\Gamma\left(\frac{k_2}{2}\right)}k_1^{k_1/2}k_2^{k_2/2}\cdot\dfrac{x^{k_1/2-1}}{(k_2+k_1x)^{(k_1+k_2)/2}}, & x>0\\ 0, & x<0\end{cases}$ k_1,k_2 正整数	$\dfrac{k_2}{k_2-2}$ $(k_2>2)$	$\dfrac{2k_2^2(k_1+k_2-2)}{k_1(k_2-2)^2(k_2-4)}$ $(k_2>4)$	
Beta-分布	$p(x)=\begin{cases}\dfrac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)}x^{p-1}(1-x)^{q-1}, & 0<x<1\\ 0, & x\leqslant 0 \text{ 或 } x\geqslant 1\end{cases}$ $p>0,q>0$ 常数	$\dfrac{p}{p+q}$	$\dfrac{pq}{(p+q)^2(p+q+1)}$	$\dfrac{\Gamma(p+q)}{\Gamma(p)}\sum_{j=0}^{\infty}\dfrac{\Gamma(p+j)(it)^j}{\Gamma(p+q+j)\Gamma(j+1)}$
对数正态分布	$p(x)=\begin{cases}\dfrac{1}{\sigma x\sqrt{2\pi}}e^{-(\ln x-a)^2/2\sigma^2}, & x>0\\ 0, & x\leqslant 0\end{cases}$ $a,\ \sigma>0$ 常数	$e^{a+\sigma^2/2}$	$e^{2a+\sigma^2}(e^{\sigma^2}-1)$	
Weibul分布	$p(x)=\begin{cases}\alpha\lambda x^{\alpha-1}e^{-\lambda x^{\alpha}}, & x>0\\ 0, & x\leqslant 0\end{cases}$ $\lambda>0,\alpha>0$ 常数	$\Gamma\left(\dfrac{1}{\alpha}+1\right)\lambda^{-1/\alpha}$	$\lambda^{-2/\alpha}\left[\Gamma\left(\dfrac{2}{\alpha}+1\right)-\left(\Gamma\left(\dfrac{1}{\alpha}+1\right)\right)^2\right]$	
拉普拉斯分布	$p(x)=\dfrac{1}{2\lambda}e^{-\frac{\lvert x-\mu\rvert}{\lambda}}$ $\lambda>0,\ \mu$ 常数	μ	$2\lambda^2$	$\dfrac{e^{i\mu t}}{1+\lambda^2t^2}$

附 录 二

普阿松分布 $P\{\xi=r\}=\frac{\lambda^r}{r!}e^{-\lambda}$ 的数值表

r	λ							
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8
0	0.904837	0.818731	0.740818	0.670320	0.606531	0.548812	0.496585	0.449329
1	0.090484	0.163746	0.222245	0.268128	0.303265	0.329287	0.347610	0.359463
2	0.004524	0.016375	0.033337	0.053626	0.075816	0.098786	0.121663	0.143785
3	0.000151	0.001092	0.003334	0.007150	0.012636	0.019757	0.028388	0.038343
4	0.000004	0.000055	0.000250	0.000715	0.001580	0.002964	0.004968	0.007669
5	—	0.000002	0.000015	0.000057	0.000158	0.000356	0.000696	0.001227
6	—	—	0.000001	0.000004	0.000013	0.000036	0.000081	0.000164
7	—	—	—	—	0.000001	0.000003	0.000008	0.000019
8	—	—	—	—	—	—	0.000001	0.000002

r	λ = 0.9	1.0	1.5	2.0	2.5	3.0	3.5	4.0
0	0.406570	0.367879	0.223130	0.135335	0.082085	0.049787	0.030197	0.018316
1	0.363913	0.367879	0.334695	0.270671	0.205212	0.149361	0.150091	0.073263
2	0.164661	0.183940	0.251021	0.270671	0.256516	0.224042	0.184959	0.146525
3	0.049398	0.061313	0.125510	0.180447	0.213763	0.224042	0.215785	0.195367
4	0.011115	0.015328	0.047067	0.090224	0.133602	0.168031	0.188812	0.195367
5	0.002001	0.003066	0.014120	0.036089	0.066801	0.100819	0.132169	0.156293
6	0.000300	0.000511	0.003530	0.012030	0.027834	0.050409	0.077098	0.104196
7	0.000039	0.000073	0.000756	0.003437	0.009941	0.021604	0.038549	0.059540
8	0.000004	0.000009	0.000142	0.000859	0.003106	0.008102	0.016865	0.029770
9	—	0.000001	0.000024	0.000191	0.000863	0.002701	0.006559	0.013231
10	—	—	0.000004	0.000038	0.000216	0.000810	0.002296	0.005292
11	—	—	—	0.000007	0.000049	0.000221	0.000730	0.001925
12	—	—	—	0.000001	0.000010	0.000055	0.000213	0.000642
13	—	—	—	—	0.000002	0.000013	0.000057	0.000197
14	—	—	—	—	—	0.000003	0.000014	0.000056
15	—	—	—	—	—	0.000001	0.000003	0.000015
16	—	—	—	—	—	—	0.000001	0.000004
17	—	—	—	—	—	—	—	0.000001

续前表

r	λ						
	4.5	5.0	6.0	7.0	8.0	9.0	10.0
0	0.011109	0.006738	0.002479	0.000912	0.000335	0.000123	0.000045
1	0.049990	0.033690	0.014873	0.006383	0.002684	0.001111	0.000454
2	0.112479	0.084224	0.044618	0.022341	0.010735	0.004998	0.002270
3	0.168718	0.140374	0.089235	0.052129	0.028626	0.014994	0.007567
4	0.189808	0.175467	0.133853	0.091226	0.057252	0.033737	0.018917
5	0.170827	0.175467	0.160623	0.127717	0.091604	0.060727	0.037833
6	0.128120	0.146223	0.160623	0.149003	0.122138	0.091090	0.063055
7	0.082363	0.104445	0.137677	0.149003	0.139587	0.117116	0.090079
8	0.046329	0.065278	0.103258	0.130377	0.139587	0.131756	0.112599
9	0.023165	0.036266	0.068838	0.101405	0.124077	0.131756	0.125110
10	0.010424	0.018133	0.041303	0.070983	0.099262	0.118580	0.125110
11	0.004264	0.008242	0.022529	0.045171	0.072190	0.097020	0.113736
12	0.001599	0.003434	0.011264	0.026350	0.048127	0.072765	0.094780
13	0.000554	0.001321	0.005199	0.014188	0.029616	0.050376	0.072908
14	0.000178	0.000472	0.002288	0.007094	0.016924	0.032384	0.052077
15	0.000053	0.000157	0.000891	0.003311	0.009026	0.019431	0.034718
16	0.000015	0.000049	0.000334	0.001448	0.004513	0.010930	0.021699
17	0.000004	0.000014	0.000118	0.000596	0.002124	0.005786	0.012764
18	0.000001	0.000004	0.000039	0.000232	0.000944	0.002893	0.007091
19	—	0.000001	0.000012	0.000085	0.000397	0.001370	0.003732
20	—	—	0.000004	0.000030	0.000159	0.000617	0.001866
21	—	—	0.000001	0.000010	0.000061	0.000264	0.000889
22	—	—	—	0.000003	0.000022	0.000108	0.000404

正态分布密度函数 $\varphi(x)=\frac{1}{\sqrt{2\pi}}e^{-x^2/2}$ 及分布函数

$\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-t^2/2}dt$ 的数值表

x	$\varphi(x)$	$\Phi(x)$	x	$\varphi(x)$	$\Phi(x)$
0.0	0.398942	0.500000	2.3	0.028327	0.989276
0.1	0.396952	0.539828	2.4	0.022395	0.991802
0.2	0.391043	0.579260	2.5	0.017528	0.993790
0.3	0.381388	0.617911	2.6	0.013583	0.995339
0.4	0.368270	0.655422	2.7	0.010421	0.996533
0.5	0.352065	0.691462	2.8	0.007915	0.997445
0.6	0.333225	0.725747	2.9	0.005953	0.998134
0.7	0.312254	0.758036	3.0	0.004432	0.998650
0.8	0.289692	0.788145	3.1	0.003267	0.999032
0.9	0.266085	0.815940	3.2	0.002384	0.999313
1.0	0.241971	0.841345	3.3	0.001723	0.999517
1.1	0.217852	0.864334	3.4	0.001232	0.999663
1.2	0.194186	0.884930	3.5	0.000873	0.999767
1.3	0.171369	0.903200	3.6	0.000612	0.999841
1.4	0.149727	0.919243	3.7	0.000425	0.999892
1.[illegible]	0.129518	0.933193	3.8	0.000292	0.999928
1.6	0.110921	0.945201	3.9	0.000199	0.999952
1.7	0.094049	0.955435	4.0	0.000134	0.999968
1.8	0.078950	0.964070	4.1	0.000089	0.999979
1.9	0.065616	0.971283	4.2	0.000059	0.999987
2.0	0.053991	0.977250	4.3	0.000039	0.999991
2.1	0.043984	0.982136	4.4	0.000025	0.999995
2.2	0.035475	0.986097	4.5	0.000016	0.999997

概率論題解

蔡 焉　楊守昌

安徽大学数学系

前　言

这本《概率论题解》，是根据复旦大学数学系编《概率论》第一册《概率论基础》（1979年版）编写的，对书中习题全部作出了解答。

本题解是在教学的基础上，参考了有关材料整理而成的。由于我们水平有限，加之编写时间仓促，错误之处在所难免，请批评指正。

我们在编写这本题解的过程中，得到了有关兄弟院校的支持和鼓励。在解题过程中曾得到了中国科技大学陈希孺教授的指导。我系刘永生付教授和杜丹青老师仔细地审阅了手稿，提出了很多修改意见。此书的编辑和出版，得到了我系概率论教研室全体同志的支持和协助。在教学过程中，我系七七级学生也提出了一些可取的意见。本书的排印工作还得到了我校印刷厂的热情支持。在此，对他们一并表示感谢。

编　者

一九八一年元月

目　　录

前　言

第一章　事件与概率 ………………………… （ 1 ）

第二章　条件概率与統計独立性 ……… （ 36 ）

第三章　随机变量与分布函数 ………… （ 66 ）

第四章　数字特徵与特徵函数 ………… （105）

第五章　极限定理 ………………………… （156）

第一章 事件与概率

本章内容 随机现象与统计独立性，样本空间与事件，古典概型，几何概率，概率空间。

1.在某城市中，共发行三种报纸 A,B,C.在这城市的居民中，订购A的占45%，订购B的占35%，订购C 的占30%，同时订购A，B 的占10%，同时订购A，C 的占 8 %，同时订购B，C的占 5 %，同时订购A，B，C 的占 3 %，试求下列百分率：(1)只订购A的；(2)只订购 A 及 B 的；(3)只订购一种报纸的；(4)正好订购两种报纸的；(5)至少订购一种报纸的；(6)不订购任何报纸的。

解：(1)P{只订购A的} $= P(A) - P\{A(B\cup C)\}$

$= P(A) - \{P(AB) + P(AC) - P(ABC)\}$

$= 0.45 - 0.10 - 0.08 + 0.03 = 0.30.$

(2)P{只订购A及B的} $= P\{(AB) - C\}$

$= P(AB) - P(ABC)$

$= 0.10 - 0.03 = 0.07$

(3)P{只订购A的} $= 0.30,$

P{只订购B的} $= P\{B - (A\cup C)\}$

$= 0.35 - (0.10 + 0.05 - 0.03) = 0.23,$

P{只订购C的} $= P\{C - (A\cup B)\}$

$= 0.30 - (0.05 + 0.08 - 0.03) = 0.20,$

$\therefore$ P{只订购一种报纸的}

$= P\{$只订购A$\} + P\{$只订购B$\} + P\{$只订购C$\}$

$= 0.30 + 0.23 + 0.20 = 0.73.$

(4) $P\{$正好订购两种报纸的$\}$

$= P\{(AB - C) \cup (AC - B) \cup (BC - A)\}$

$= P(AB - ABC) + P(AC - ABC) +$

$+ P(BC - ABC)$

$= (0.1 - 0.03) + (0.08 - 0.03) + (0.05$

$- 0.03) = 0.07 + 0.05 + 0.02 = 0.14.$

(5) $P\{$至少订购一种报纸的$\} = P\{$只订一种的$\}$

$+ P\{$恰订两种的$\} + P\{$恰订三种的$\}$

$= 0.73 + 0.14 + 0.03 = 0.90.$

(6) $P\{$不订任何报纸的$\} = 1 - 0.90 = 0.10.$

2．若A，B，C 是随机事件，说明下列关系式的概率意义：(1) $ABC = A$； (2) $A \cup B \cup C = A$； (3) $AB \subset C$；(4) $A \subset \overline{BC}$。

解： (1) $ABC = A \Leftrightarrow BC \supset A$ ($ABC \subset A$显然) $\Leftrightarrow B \supset A$且$C \supset A$，若A发生，则B与C必同时发生。

(2) $A \cup B \cup C = A \Leftrightarrow B \cup C \subset A \Leftrightarrow B \subset A$且$C \subset A$，B发生或C发生，均导致A发生。

(3) $AB \subset C \Leftrightarrow$ A与B同时发生必导致C发生。

(4) $A \subset \overline{BC} \Leftrightarrow A \subset \overline{B} \cup \overline{C}$，A发生，则B与C至少有一不发生。

3．试把$A_1 \cup A_2 \cup \cdots \cup A_n$ 表示成 n 个两两互不相容事件的和。

解： $A_1 \cup A_2 \cup \cdots \cup A_n$

$= A_1 + (A_2 - A_1) + \cdots$

$+ (A_n - A_1 - \cdots - A_{n-1})$

（或）$= A_1 + A_2\overline{A}_1 + \cdots + A_n\overline{A}_1\overline{A}_2\cdots\overline{A}_{n-1}$.

4．在某班学生中任选一个同学，以事件A表示选到的是男同学，事件B表示选到的人不喜欢唱歌，事件C表示选到的人是运动员。(1)表述 $AB\overline{C}$ 及 $A\overline{B}C$；(2)什么条件下成立 $ABC = A$；(3)何时成立 $\overline{C}\subset B$；(4)何时同时成立 $A = B$ 及 $\overline{A} = C$。

解：(1)$AB\overline{C}$ = {抽到的是男同学，又不爱唱歌，又不是运动员}； $A\overline{B}C$ = {抽到的是男同学，又爱唱歌，又是运动员}。

(2)$ABC = A \Leftrightarrow BC \supset A$，当男同学都不爱唱歌且都是运动员时成立。

(3)当不是运动员的学生必是不爱唱歌的时，$\overline{C}\subset B$ 成立。

(4)$A = B$ 及 $\overline{A} = C \Leftrightarrow A = B = \overline{C}$，当男学生的全体也就是不爱唱歌的学生全体，也就不是运动员的学生全体时成立。也可以表述为：当男学生不爱唱歌且不爱唱歌的一定是男学生，并且男学生不是运动员且不是运动员的是男学生时成立。

5．用模球模型造一例，指出样本空间及各种事件运算。

解：设袋中有三个球，编号为1，2，3，每次模一个球。样本空间共有3个样本点（1），（2），（3）。设 $A = \{1,2\}$，$B = \{1, 3\}$，$C = \{3\}$，则 $\overline{A} = \{3\}$，$A\cup B = \{1,2,3\}$，$A\cap B = \{1\}$，$A - B = \{2\}$，$A + C = \{1,2,3\}$。

6．若 A，B，C，D 是四个事件，试用这四个事件表示下列各事件：(1)这四个事件至少发生一个；(2)这四个事件恰好发生两个；(3)A，B都发生而C，D都不发生；(4)这四

个事件都不发生；(5)这四个事件中至多发生一个。

解：(1){至少发生一个}$=A\cup B\cup C\cup D$。

(2){恰发生两个}$=AB\bar{C}\bar{D}+AC\bar{B}\bar{D}+AD\bar{B}\bar{C}$
$+BC\bar{A}\bar{D}+CD\bar{A}\bar{B}+BD\bar{A}\bar{C}$。

(3){A,B都发生而C,D都不发生}$=AB\bar{C}\bar{D}$。

(4){都不发生}$=\bar{A}\bar{B}\bar{C}\bar{D}=\overline{A\cup B\cup C\cup D}$。

(5){至多发生一个}$=\bar{A}\bar{B}\bar{C}\bar{D}+A\bar{B}\bar{C}\bar{D}$
$+B\bar{A}\bar{C}\bar{D}+C\bar{A}\bar{B}\bar{D}+D\bar{A}\bar{B}\bar{C}$
$=\overline{AB\cup AC\cup AD\cup BC\cup BD\cup CD}$。

*7. 从0，1，2，…，9中随机地取出5个数(可重复)，以E_i记某些数正好出现i次这一事件(例如52353，既属于E_1，也属于E_2及E_0)，试用文图表示$E_0,E_1,\cdots,E_6$的关系。

解：分析一下E_i之间的关系。先依次设样本点$\omega\in E_i$，再分析此ω是否属于$E_j(j\neq i)$，$E_jE_K(j\neq i,\ k\neq i)$等。(1)$E_6$为不可能事件。(2)若$\omega\in E_5$，则$\omega\bar{\in}E_i(i=1,2,3,4)$，即$E_5E_i=\phi$。(3)若$\omega\in E_4$，则必有$\omega\in E_1$，于是$E_4\subset E_1$。若$\omega\in E_4$，则$\omega\bar{\in}E_2$，$\omega\bar{\in}E_3$。(4)若$\omega\in E_3$，则必有$\omega\in E_2$或$\omega\in E_1$之一发生，但$\omega\bar{\in}E_1E_2$。由此得$E_3E_1\cup E_3E_2=E_3$，$E_1E_2E_3=\phi$。(5)若$\omega\in E_2$，则必有$\omega\in E_1$或$\omega\in E_3$之一发生，由此得$E_2E_1\cup E_2E_3=E_2$。(6)$E_1$中还有这样的点$\omega$：12345，它仅属于$E_1$，而不再属于其它$E_i(i\neq1,0)$。诸$E_i$之间的关系用文图表示如图1。

<table>
<tr><td>E_1</td><td>E_1E_4</td><td rowspan="3">E_5</td></tr>
<tr><td>E_1E_2</td><td>E_1E_3</td></tr>
<tr><td colspan="2">E_2E_3</td></tr>
</table>

$E_6=\phi$，$E_0=\Omega$

图 1

8．证明下列等式：

(1) $C_n^1+2C_n^2+3C_n^3+\cdots+nC_n^n=n2^{n-1}$；

(2) $C_n^1-2C_n^2+3C_n^3-\cdots+(-1)^{n-1}nC_n^n=0$；

(3) $\sum_{k=0}^{a-r}C_a^{k+r}C_b^k=C_{a+b}^{a-r}$。

证：(1)因为

$$(1+x)^n=1+C_n^1x+C_n^2x^2+\cdots+C_n^nx^n$$

两边对x求导得

$$n(1+x)^{n-1}=C_n^1+2C_n^2x+\cdots+nC_n^nx^{n-1},$$

在其中令$x=1$即得所欲证。

(2)在上式中令$x=-1$即得所欲证。

(3)要原式有意义，必须 $0\leqslant r\leqslant a$。由于

$C_{a+b}^{a-r}=C_{a+b}^{b+r}$，$C_b^k=C_b^{b-k}$，此题即等于要证

$$\sum_{k=0}^{a}C_a^{k+r}C_b^{b-k}=C_{a+b}^{b+r},\quad 0\leqslant r\leqslant a.$$

利用幂级数乘法可证明此式。因为

$$(x+1)^a(x+1)^b=(x+1)^{a+b},$$

比较等式两边x^{b+r}的系数即得证。

9．袋中有白球5只，黑球6只，陆续取出三球，求顺序为黑白黑的概率。

解：$p=A_6^1A_5^1A_5^1/A_{11}^3=\dfrac{5}{33}\doteq0.15$。

10．一部五本头的文集，按任意次序放到书架上去，试求下列概率：(1)第一卷出现在旁边；(2)第一卷及第五卷出现在旁边；(3)第一卷或第五卷出现在旁边；(4)第一卷及第五卷都不出现在旁边；(5)第三卷正好在正中。

解：（1）第一卷出现在旁边，可能出现在左边或右边，剩下四卷可在剩下四个位置上任意排，所以

$$p = 2 \times 4!/5! = 2/5$$

（2）可能有第一卷出现在左边而第五卷出现在右边，或者第一卷出现在右边而第五卷出现在左边，剩下三卷可在中间三个位置上任意排，所以

$$p = 2 \times 3!/5! = 1/10.$$

（3）$p = P$｛第一卷出现在旁边｝$+ P$｛第五卷出现在旁边｝$- P$｛第一卷及第五卷出现在旁边｝

$$= \frac{2}{5} + \frac{2}{5} - \frac{1}{10} = \frac{7}{10}.$$

（4）这里事件是（3）中事件的对立事件，所以

$$p = 1 - 7/10 = 3/10.$$

（5）第三卷居中，其余四卷在剩下四个位置上可任意排，所以

$$p = 1 \times 4!/5! = 1/5.$$

11. 把1，2，3，4，5诸数各写在一小纸片上，任取其三而排成自左向右的次序，求所得数是偶数的概率。

解： 末位数只可能是 2 或 4 。当末位数是 2（或 4）时，前两位数字从剩下四个数字中选排，所以

$$p = 2 \times A_4^2 / A_5^3 = 2/5.$$

12. 在一个装有n只白球，n只黑球，n只红球的袋中，任取m只球，求其中白、黑、红球分别有m_1, m_2, m_3（$m_1 + m_2 + m_3 = m$）只的概率。

解： $p = C_n^{m_1} C_n^{m_2} C_n^{m_3} / C_{3n}^m.$

13. 甲袋中有 3 只白球，7 只红球，15只黑球，乙袋中

有10只白球，6只红球，9只黑球，现从两袋中各取一球，求两球颜色相同的概率。

解： P{两球颜色相同} = P{两球均白} + P{两球均黑} + P{两球均红}

$$= \frac{3}{25}\times\frac{10}{25}+\frac{7}{25}\times\frac{6}{25}+\frac{15}{25}\times\frac{9}{25}=\frac{207}{625}=0.33.$$

14．由盛有号码1,2,…，N的球的箱子中有放回地摸了n次球，依次记下其号码，试求这些号码按严格上升次序排列的概率。

解： 若取出的号码是按严格上升次序排列，则 n个号码必然全不相同，$n\leqslant N$．n 个不同号码可产生 n! 种不同的排列，其中只有一个是按严格上升次序的排列，也就是说，一种组合对应一种严格上升排列，所以共有C_N^n 种按严格上升次序的排列．总可能场合数为N^n，故题中欲求的概率为

$$p = C_N^n/N^n.$$

*15．在上题中这些号码按上升（不一定严格）次序排列的概率。

解法一： 先引入重复组合的概念．从n 个不同的元素里，每次取出m个元素，元素可以重复选取，不管怎样的顺序并成一组，叫做从 n 个元素里每次取m个元素的重复组合，其组合种数记为$\widetilde{C}_n^m$．重复组合种数与不重复（通常）组合种数之间有如下关系式：

$$\widetilde{C}_n^m = C_{n+m-1}^m.$$

这个公式的证明思路是，把n个不同的元素编号为 1,2,…，n，再把重复组合的每一组中数从小到大排列，每个数依次加上0,1,…，m－1，则这一组数就变成了从1,2,…，n＋m

-1共$n+m-1$个数中，取出m个数的不重复组合中的一组。这种运算构成两者之间一一对应。

若取出n个号码按上升（不一定严格）次序排列，与上题同理可得，一个重复组合对应一种按上升次序的排列，所以共有$\widetilde{C}_N^n$种按上升次序的排列，总可能场合数为N^n，从而

$$p=\widetilde{C}_N^n/N^n=C_{N+n-1}^n/N^n$$

解法二：现按另一思路求解。取出的n个数中间可设$n-1$个间壁。当取出的n个数全部相同时，可以看成中间没有间壁，故间壁有C_{n-1}^0种取法；这时只需取一个数字，有C_N^1种取法；这种场合的种数有$C_{n-1}^0C_N^1$种。当n个数由两样数构成时，看成间壁数为1，分成左右两段，分别由小大两个数填上，而间壁的位置有C_{n-1}^1种取法；数字有C_N^2种取法；这种场合的种数有$C_{n-1}^1C_N^2$种。当n个数由三样数构成时，可得场合种数为$C_{n-1}^2C_N^3$种，等等。最后，当n个数均为不同数字时，有$n-1$个间壁，有C_{n-1}^{n-1}种取法；数字有C_N^n种取法；这种场合种数有$C_{n-1}^{n-1}C_N^n$种。所以共有有利场合数为：

$$m_1=C_{n-1}^0C_N^1+C_{n-1}^1C_N^2+C_{n-1}^2C_N^3+\cdots+C_{n-1}^{n-1}C_N^n=C_{N+n-1}^n.$$

此式证明见本章第8题（3）。总可能场合数为$n_1=N^n$，故所求的概率为

$$p=m_1/n_1=C_{N+n-1}^n/N^n.$$

16．任意从数列1，2，…，N中不放回地取出n个数并按大小排列成：$x_1<x_2<\cdots<x_m<\cdots<x_n$，试求$x_m=M$的概

率，这里$1\leqslant M\leqslant N$。

解：因为不放回，所以 n 个数不重复。从$\{1, 2, \cdots, M-1\}$中取出$m-1$个数，从$\{M+1, \cdots N\}$中取出$n-m$个数，数M一定取出，把这n个数按大小次序重新排列，则必有 $x_m = M$。故

$$p = C_{M-1}^{m-1}\ C_1^1\ C_{N-M}^{n-m}\ /\ C_N^n\ .$$

当$M-1<m-1$或$N-M<n-m$时，概率$p=0$。

*17．上题中，若采用有放回取数，这时 $x_1\leqslant x_2\leqslant\cdots\leqslant x_n$，试求$x_m=M$的概率。

解：从$1, 2, \cdots, N$ 中有放回地取 n 个数，这 n 个数有三类：$<M$，$=M$，$>M$。如果我们固定 k_1次是取到$<M$的数，k_2次是取到$>M$的数，当然其余一定是取到M的。

当次数固定后，$<M$的有$(M-1)^{k_1}$种可能的取法(因为每一次都可以从$M-1$个数中取一个)，$>M$ 的有$(N-M)^{k_2}$种可能的取法，而$=M$的只有一种取法(即全是M)，所以可能的取法有 $(M-1)^{k_1}\ (N-M)^{k_2}$ 种。对于确定的 k_1，k_2来说，在 n 次取数中，固定哪 k_1次取到$<M$的数，哪 k_2次取到$>M$的数，这共有$C_n^{k_1}\times C_{n-k_1}^{k_2}$种不同的固定方式，因此 k_1 次取到$<M$的数，k_2次取到$>M$ 的数的可能取法有

$$C_n^{k_1}C_{n-k_1}^{k_2}\ (M-1)^{k_1}\ (N-M)^{k_2}\text{种}.$$

设B表示事件“把取出的 n 个数从小到大重新排列后第m个数等于M”，则B出现就是k_1次取到$<M$的数，k_2 次取到$>M$的数，$0\leqslant k_1\leqslant m-1$，$0\leqslant k_2\leqslant n-m$，因此B 包含的所有可能的取法有

$$\sum_{k_1=0}^{m-1}\sum_{k_2=0}^{n-m}C_n^{k_1}C_{n-k_1}^{k_2}\quad(M-1)^{k_1}\quad(N-M)^{k_2}\quad 种。$$

所以

$$P(B)=\frac{1}{N^n}\sum_{k_1=0}^{m-1}\sum_{k_2=0}^{n-m}C_n^{k_1}\ C_{n-k_1}^{k_2}\quad\times$$

$$\times(M-1)^{k_1}\ (N-M)^{k_2}.$$

18．从 6 双不同的手套中任取 4 只，问其中恰有一双配对的概率是多少？

解：有利场合是，先从 6 双中取出一双，其两只全取出；再从剩下的 5 双中取出两双，从其每双中取出一只。所以欲求的概率为

$$p=C_6^1C_2^2C_5^2C_2^1C_2^1\ /\ C_{12}^4=\frac{16}{33}=0.48.$$

19．从 n 双不同的鞋子中任取2r(2r<n)只，求下列事件发生的概率：(1)没有成对的鞋子；(2)只有一对鞋子；(3)恰有两对鞋子；(4)有r对鞋子。

解：(1)有利场合是，先从n双中取出2r双，再从每双中取出一只。

$$p=C_n^{2r}(C_2^1\ )^{2r}/C_{2n}^{2r},\quad(2r<n).$$

(2)有利场合是，先从n双中取出一双，其两只全取出，再从剩下的n－1双中取出2r－2双，从其每双中取出一只。

$$p=C_n^1C_2^2C_{n-1}^{2r-2}\ (C_2^1\ )^{2r-2}/C_{2n}^{2r}$$

$$=n2^{2r-2}C_{n-1}^{2r-2}\ /C_{2n}^{2r}.$$

(3)$p=2^{2r-4}\ C_n^2\ C_{n-2}^{2r-4}\ /\ C_{2n}^{2r}.$

(4)$p = C_n^r (C_2^2)^r / C_{2n}^{2r} = C_n^r / C_{2n}^{2r}$。

20．袋中有 n 只球，记有号码 1,2,…,n ，求下列事件的概率：(1)任意取出两球，号码为1,2；(2)任意取出3球，没有号码1；(3)任意取出5球，号码1,2,3中至少出现一个。

解：(1)P{任意取两球，号码为1,2} = $1/C_n^2$。

(2)任取 3 个球无号码 1 ，有利场合是从除去 1 号球外的n－1个球中任取 3 个球的组合数，故

$$P\{任取3球，无号码1\} = C_{n-1}^3 / C_n^3。$$

(3)P{任取 5 球，号码1,2,3中至少出现 1 个}

$=1-P\{任取5球，号码1,2,3不出现\}$

$= 1 - C_{n-3}^5 / C_n^5$。

其中任取 5 球无号码1,2,3，有利场合是从除去1,2,3号球外的n－ 3 个球中任取 5 个球的组合数。

21．袋中装有1,2,…，N 号的球各一只，采用（1）有放回；（2）不放回方式摸球，试求在第 k 次摸球时首次摸到 1 号球的概率。

解：(1)有利场合是，前k－1次从N－1个号中（除 1 号外）抽取，第k次取到 1 号球，

$$p = (N-1)^{k-1} \cdot 1/N^k = (N-1)^{k-1}/N^k。$$

(2)考虑前k次摸球的情况，

$$p = A_{N-1}^{k-1} \cdot 1 / A_N^k = 1/N。$$

22．甲有n＋ 1 个硬币，乙有n个硬币，双方投掷之后进行比较，求甲掷出的正面比乙掷出的正面多的概率。

解法一：设A = {甲掷出正面数＞乙掷出正面数}，

B = {甲掷出反面数＞乙掷出反面数}。

考虑$\overline{A}$={甲掷出正面数≤乙掷出正面数}。设$\overline{A}$发生。若乙掷出n次正面，则甲至多掷出n次正面，也就是说乙掷出0次反面，甲至少掷出1次反面，从而甲掷出反面数>乙掷出反面数。若乙掷出n－1次正面，则甲至多掷出n－1次正面，也就是说乙掷出1次反面，甲至少掷出2次反面，从而也有甲掷出反面数>乙掷出反面数，等等。由此可得

$\overline{A}$={甲掷出正面数≤乙掷出正面数}

={甲掷出反面数>乙掷出反面数}=B。

$$\therefore P(A)+P(B)=P(A)+P(\overline{A})=1.$$

显然A与B是等可能的，因为每人各自掷出正面与反面的可能性相同，所以P(A)=P(B)，从而$P(A)=\frac{1}{2}$。

解法二：甲掷n+1个硬币共有2^{n+1}个等可能场合，其中有C_{n+1}^{0}个出现0次正面，有C_{n+1}^{1}个出现1次正面，…，C_{n+1}^{n+1}个出现n+1次正面。乙掷n个硬币共有2^{n}个等可能场合，其中有C_{n}^{0}个出现0次正面，C_{n}^{1}个出现1次正面，…，C_{n}^{n}个出现n次正面。若甲掷n+1个硬币，乙掷n个硬币，则共有$n_1=2^{n+1}\cdot 2^{n}=2^{2n+1}$种等可能场合，其中甲掷出正面比乙掷出正面多的有利场合数有

$$\begin{aligned} m_1 = & C_{n+1}^{1}C_{n}^{0}+C_{n+1}^{2}(C_{n}^{0}+C_{n}^{1})+ \\ & +C_{n+1}^{3}(C_{n}^{0}+C_{n}^{1}+C_{n}^{2})+\cdots \\ & +C_{n+1}^{n}(C_{n}^{0}+C_{n}^{1}+\cdots+C_{n}^{n-1})+ \\ & +C_{n+1}^{n+1}(C_{n}^{0}+C_{n}^{1}+\cdots+C_{n}^{n}). \end{aligned}$$

利用公式　$C_{n+1}^{r}=C_{n}^{r}+C_{n}^{r-1}$及$C_{n+1}^{n+1}=C_{n}^{n}$得

$$m_1=(C_n^0+C_n^1)C_n^0+(C_n^1+C_n^2)(C_n^0+C_n^1)+$$

$$+(C_n^2+C_n^3)(C_n^0+C_n^1+C_n^2)+$$

$$\cdots+(C_n^{n-1}+C_n^n)(C_n^0+C_n^1+\cdots+C_n^{n-1})$$

$$+C_n^n(C_n^0+C_n^1+\cdots+C_n^n)$$

$$=\left[(C_n^0)^2+C_n^1C_n^0\right]+\left[(C_n^1)^2+C_n^1C_n^0+\right.$$

$$\left.+C_n^2\sum_{i<2}C_n^i\right]+\left[(C_n^2)^2+C_n^2\sum_{i<2}C_n^i+\right.$$

$$\left.+C_n^3\sum_{i<3}C_n^i\right]+\cdots+\left[(C_n^{n-1})^2+\right.$$

$$\left.+C_n^{n-1}\sum_{i<n-1}C_n^i+C_n^n(\sum_{i<n}C_n^i\right]+$$

$$+\left((C_n^n)^2+C_n^n\sum_{i<n}C_n^i\right)$$

$$=\sum_{i=0}^n(C_n^i)^2+2\sum_{n\geq j>i\geq 0}C_n^iC_n^j=\left(\sum_{i=0}^nC_n^i\right)^2$$

$$=(2^n)^2=2^{2n}.$$

所以欲求的概率为

$$p=m_1/n_1=2^{2n}/2^{2n+1}=\frac{1}{2}.$$

应注意，甲掷出$0,1,\cdots,n+1$个正面的$n+2$个场合不是等可能的。

23.（De Méré 问题）一颗骰子投4次至少得到一个六点，与两颗骰子投24次至少得到一个双六这两件事，哪一个有更多的机会遇到？

解： 事件"一颗投4次至少得一个六点"的对立事件为"一颗投4次没有一个六点"，后者的有利场合数为，除去六点外的剩下五个点允许重复地排在四个位置上的排列数，故

$$P\{\text{一颗投4次至少得一个六点}\}$$
$$=1-P\{\text{一颗投4次没有一个六点}\}$$
$$=1-5^4/6^4=0.5177.$$

投两颗骰子共有36种可能结果，除双六（6,6）点外，还有35种结果，故

$$P\{\text{两颗投24次至少得一个双六}\}$$
$$=1-P\{\text{两颗投24次没有一个双六}\}$$
$$=1-35^{24}/36^{24}=0.4914.$$

比较知，前者机会较大。

24.从52张扑克牌中任意抽取13张来，问有5张黑桃，3张红心，3张方块，2张草花的概率。

解： $p=C_{13}^5 C_{13}^3 C_{13}^3 C_{13}^2 / C_{52}^{13}=0.0129$.

25.桥牌游戏中（四人各从52张纸牌中分得13张），求4张A集中在一个人手中的概率。

解： $$p=\frac{C_4^1 C_4^4 C_{48}^9 C_{39}^{13} C_{26}^{13} C_{13}^{13}}{C_{52}^{13} C_{39}^{13} C_{26}^{13} C_{13}^{13}}=\frac{4\times C_{48}^9}{C_{52}^{13}}$$

$$=0.0106.$$

或解为，4张A集中在特定一人手中的概率为 $C_4^4 C_{48}^9/C_{52}^{13}$，所以4张A集中在一个人手中的概率为

$$p=4\times C_{48}^9 / C_{52}^{13}=0.0106 .$$

26.在扑克牌游戏中（从52张牌中任取5张），求下列事件的概率：（1）以A打头的同花顺次五张牌；（2）其

它同花顺次五张牌；（3）有四张牌同点数；（4）三张同点数且另两张也同点数；（5）五张同花；（6）异花顺次五张牌；（7）三张同点数，另外两张不同点数；（8）五张中有两对；（9）五张中有一对；（10）共它情况。

解：(1) $p=4/C_{52}^5=0.0000015$。

这里设A只打大头，若认为A可打两头AKQJ10及 A2345，则答案有变，下同。

(2)取出的五张可以由K,Q,…，6 八个数中之一打头，所以

$$p=C_4^1C_8^1/C_{52}^5=0.0000123.$$

(3)取出的四张同点牌为13个点中某一点，再从剩下48张牌中取出1张，所以

$$p=C_{13}^1C_4^4C_{48}^1/C_{52}^5=0.00024.$$

(4)取出的3张同点占有13个点中一个点，接着取出的两张同点占有其余12个点中的一个点，所以

$$p=C_{13}^1C_4^3C_{12}^1C_4^2/C_{52}^5=0.00144.$$

(5) 5张同花可以是四种花中任一种，在同一种花中，5张牌占有13个点中5个点，所以

$$p=C_4^1C_{13}^5/C_{52}^5=0.00198$$

(6){异花顺次五张牌}={顺次五张牌}-{同花顺次五张牌}。顺次五张牌分别以 A，K，…，6 九个数中之一打头，每张可以有四种不同的花；而同花顺次中花色只能是四种花中一种。所以

$$p=P\{\text{顺次五张牌}\}-P\{\text{同花顺次五张牌}\}$$
$$=[C_9^1(C_4^1)^5-C_4^1C_9^1]/C_{52}^5=0.0000294.$$

(7)三张同点牌占有13个点中一个点，另外两张不同点占有剩下12个点中两个点，所以

$$p = C_{13}^1 C_4^3 C_{12}^2 (C_4^1)^2 / C_{52}^5 = 0.0211 .$$

(8) $P\{$五张中有两对$\} = P\{$五张中两对不同点$\} +$
$+ P\{$五张中两对同点$\}$

$$= C_{13}^2 C_4^2 C_4^2 C_{11}^1 C_4^1 / C_{52}^5 +$$

$$+ C_{13}^1 C_4^4 C_{12}^1 C_4^1 / C_{52}^5 = 0.0475.$$

(9) $p = C_{13}^1 C_4^2 C_{12}^3 (C_4^1)^3 / C_{52}^5 = 0.423.$

(10)若记（i）中事件为A_i，则$A_1 \subset A_5$，$A_2 \subset A_5$，$A_3 \subset A_8$，$A_4 \subset A_9$ 而事件A_5，…，A_9两两不相容，所以

$$p = 1 - P(\bigcup_{i=5}^{9} A_i) = 1 - \sum_{i=5}^{9} P(A_i) = 0.506.$$

27．某码头只能容纳一只船，现预知某日将独立来到两只船，且在24小时内各时刻来到的可能性都相等，如果它们需要停靠的时间分别为 3 小时及 4 小时，试求有一船要在江中等待的概率．

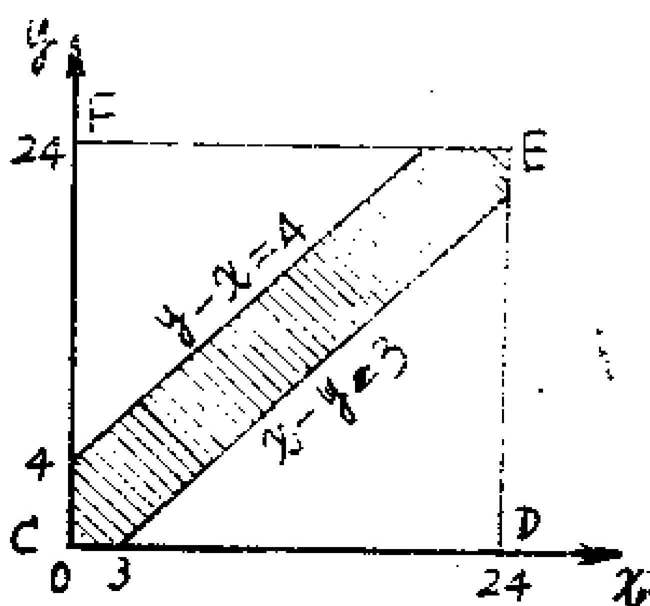


图 2

解：设 x，y 分别为此二船到达码头的时间，则

$$0 \leqslant x \leqslant 24, \quad 0 \leqslant y \leqslant 24.$$

两船到达码头的时间与由上述条件决定的正方形内的点是一一对应的（图2）．

设A表事件“一船要等待空出码头”，则A发生意味着同时满足下述两不等式

$$x-y\leqslant 3,\ y-x\leqslant 4.$$

由几何概率得，事件A的概率，等于正方形CDEF中直线$x-y\leqslant 3$及$y-x\leqslant 4$之间的部分的面积，与正方形CDEF的面积之比，即

$$P(A)=[24^2-(\tfrac{1}{2}\times 20^2+\tfrac{1}{2}\times 21^2)]/24^2$$
$$=311/1152\approx 0.27.$$

28．两人约定于7点到8点在某地会面，试求一人要等另一人半小时以上的概率。

解： 设x，y分别为此二人到达时间，则$7\leqslant x\leqslant 8$，$7\leqslant y\leqslant 8$．显然，此二人到达时间(x,y)与由上述条件决定的正方形CDEF内的点是一一对应的（图3）。

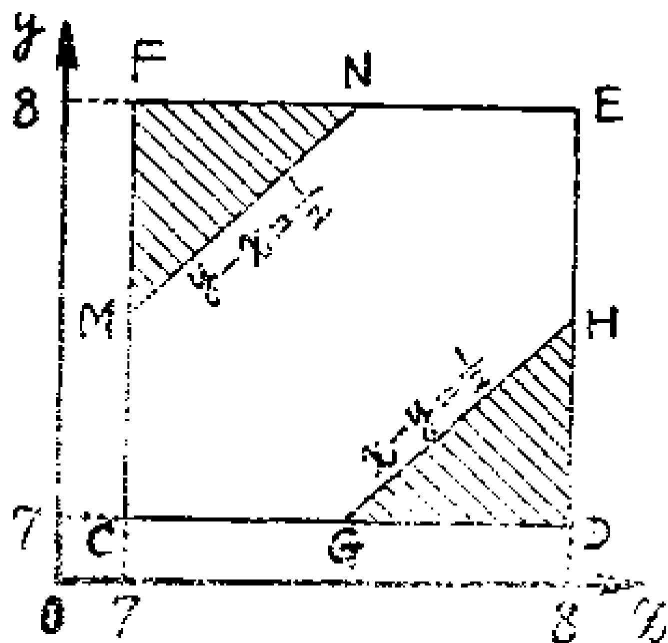


图 3

设A表事件“其中一人必须等另外一人的时间$\frac{1}{2}$小时以上”，则A发生意味着满足如下不等式

$$x-y>\tfrac{1}{2}\text{或}y-x>\tfrac{1}{2}.$$

由几何概率得，事件A的概率，等于ΔGDH及ΔFMN的面积之和与正方形CDEF的面积之比，所以

$$P(A)=\tfrac{1}{2}(\tfrac{1}{2}\times\tfrac{1}{2}+\tfrac{1}{2}\times\tfrac{1}{2})/(1\times 1)=\tfrac{1}{4}.$$

29. 在一线段AB中随机地取两个点X_1与X_2，求AX_1，X_1X_2，X_2B可以构成一个三角形的概率。（三线段能构成三角形的充要条件是任意两边之和大于第三边）

解：设$AB=a$，$AX_1=x_1$，$AX_2=x_2$，则

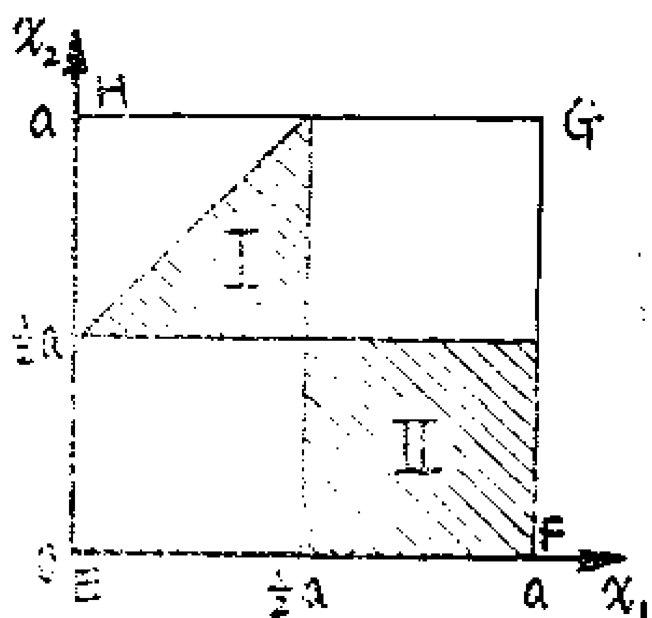


图 4

$$0\leqslant x_1\leqslant a,$$
$$0\leqslant x_2\leqslant a,$$

(x_1,x_2)与由上述条件决定的正方形EFGH内的点是一一对应的（图4）。

（Ⅰ）设 $x_2>x_1$。$AX_1=x_1$，$X_1X_2=x_2-x_1$，$X_2B=a-x_2$，则三线段构成三角形的充要条件是

$$\begin{cases} x_1+(x_2-x_1)>a-x_2\Leftrightarrow x_2>\frac{1}{2}a \\ x_1+(a-x_2)>(x_2-x_1)\Leftrightarrow x_1+\frac{1}{2}a>x_2 \\ (x_2-x_1)+(a-x_2)>x_1\Leftrightarrow x_1<\frac{1}{2}a, \end{cases}$$

这决定三角形区域Ⅰ（图4）。

（Ⅱ）设 $x_1>x_2$。$AX_1=x_1$，$X_1X_2=x_1-x_2$，$X_2B=a-x_2$，则三线段构成三角形的充要条件是

$$\begin{cases} x_1+(x_1-x_2)>a-x_2\Leftrightarrow x_1>\frac{1}{2}a \\ (x_1-x_2)+(a-x_2)>x_1\Leftrightarrow x_2<\frac{1}{2}a \\ x_1+(a-x_2)>x_1-x_2\Leftrightarrow a>0, \end{cases}$$

这决定矩形区域Ⅱ.

（Ⅱ）当$x_1 = x_2$时，不能构成三角形。

由几何概率知，

$$P\{三线段构成三角形\} = \frac{\Delta(\text{I})面积 + 矩形(\text{II})面积}{正方形EFGH面积}$$

$$= \left(\frac{1}{2} \times \frac{1}{2}a \times \frac{1}{2}a + \frac{1}{2}a \times \frac{1}{2}a\right)/a^2 = \frac{3}{8}$$

30．在线段〔0，1〕上任意投三个点，问由0至三点的三线段，能构成三角形与不能构成三角形这两个事件中哪一个事件的概率大。

解： 设0到三点的三线段长分别为x, y, z，即相应的右端点坐标为 x, y, z，显然有 $0 \leqslant x, y, z \leqslant 1$。这三条线段构成三角形的充要条件是： $x + y > z, x + z > y, y + z > x$。

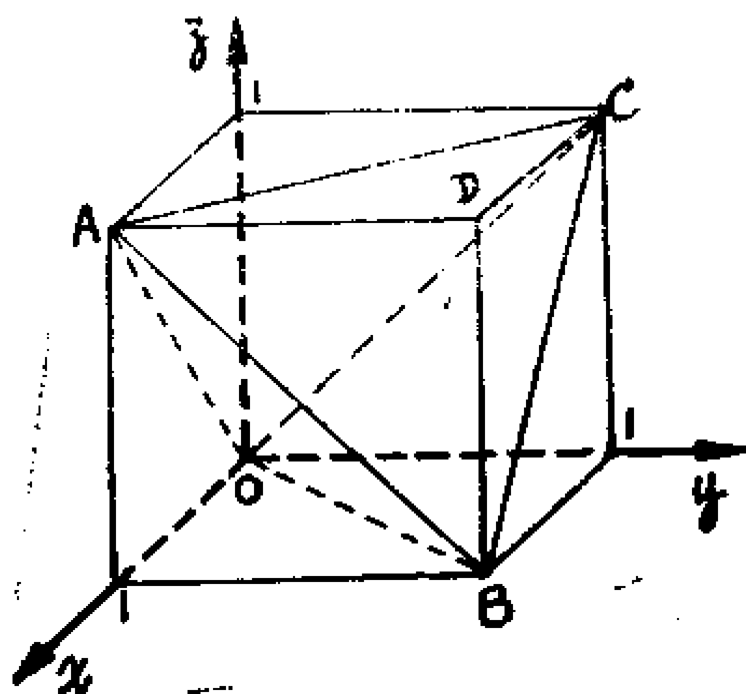


图 5

在线段$[0,1]$上任意投三点 x,y,z，与立方体$0\leqslant x\leqslant 1$，$0\leqslant y\leqslant 1, 0\leqslant z\leqslant 1$中的点$(x,y,z)$一一对应，可见所求“构成三角形”的概率，等价于在边长为1的立方体T 中均匀地掷点，而点落在 $x+y>z, x+z>y, y+z>x$ 区域中的概率；这也就是落在图5中由$\Delta ADC, \Delta ADB, \Delta BDC, \Delta AOC, \Delta AOB$，$\Delta BOC$所围成的区域G中的概率。由于 $V(T)=1$，$V(G)=1^3-3\times\frac{1}{3}\times\frac{1}{2}\times 1^3=\frac{1}{2}$，

$$\therefore p=V(G)/V(T)=\frac{1}{2}.$$

由此得，能与不能构成三角形两事件的概率一样大。

31. 在一张打上方格的纸上投一枚直径为1的硬币，方格要多小才能使硬币与线不相交的概率小于1%。

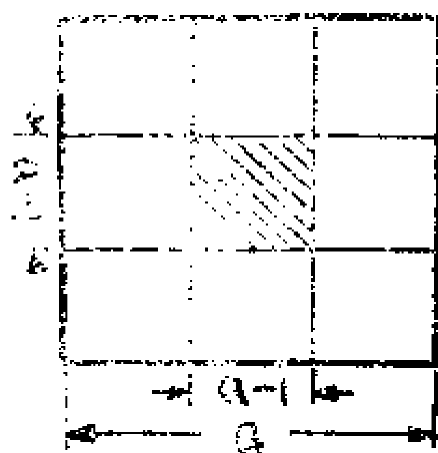


图 6

解： 设方格边长为a。当硬币圆心落于图6中阴影部分才与边界不相交（图中只取一个方格）。由几何概率得

$$P\{\text{硬币与线不相交}\}=\frac{\text{阴影部分面积}}{\text{方格面积}}$$

$$=(a-1)^2/a^2.$$

令 $$(a-1)^2/a^2=0.01$$

因为当$a\leqslant 1$时，硬币必与线相交（必然事件），故只需考虑$a>1$。由上式得

$$(a-1)/a=0.1, \qquad a=1\tfrac{1}{9}.$$

即当方格边长$a<1\frac{1}{9}$时，才能使硬币与线不相交的概率小于1%。

32．从(0，1)中随机地取两个数，求下列概率：(1)两数之和小于1.2；(2)两数之积小于$\frac{1}{4}$；(3)以上两条件同时满足。

解：从(0,1)中取出的两数分别记为x，y，则(x，y)与正方形ABCD内的点一一对应(如图)。

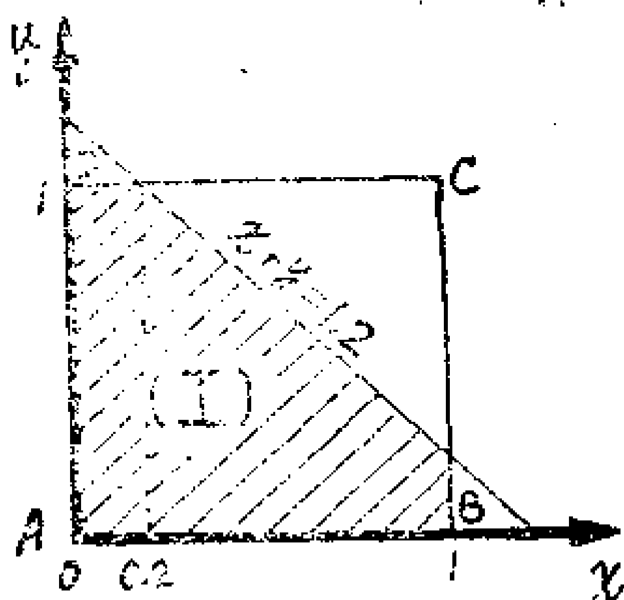


图　7

(1)直线$x+y=1.2$与BC交点坐标为(1，0.2)，与DC交点坐标为(0.2，1)，所以由几何概率可得

$$P\{\text{两数之和小于}1.2\}=\frac{\text{阴影区域(I)面积}}{\text{正方形面积}}$$

$$=(1-\frac{1}{2}\times0.8\times0.8)/1=0.68,$$

(2)双曲线$xy=\frac{1}{4}$与BC交点坐标为$(1,\frac{1}{4})$(图8)与DC交点坐标为$(\frac{1}{4},1)$，所以由几何概率得

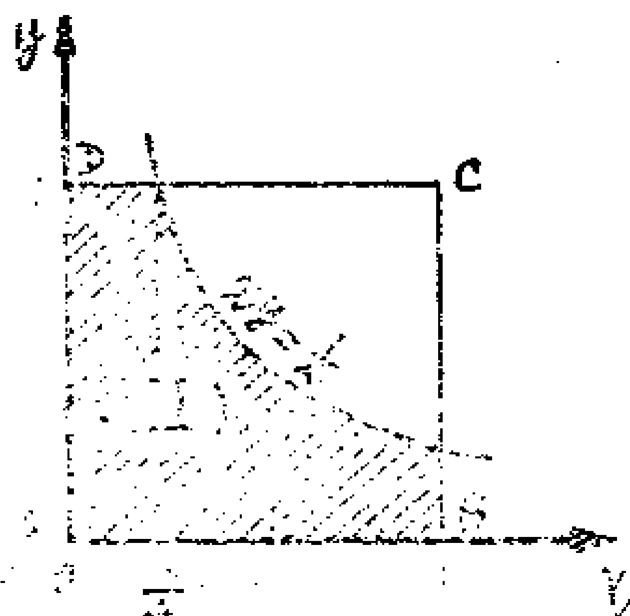


图　8

$$P\{\text{两数之积小于}\frac{1}{4}\}$$

$$=\frac{\text{阴影区域(I)面积}}{\text{正方形面积}}$$

$$=\frac{1}{4}\times1+\int_{\frac{1}{4}}^{1}\frac{1}{4x}dx$$

$$=\frac{1}{4}+\frac{1}{4}\ln x\Big|_{\frac{1}{4}}^{1}$$

$$=\frac{1}{4}+\frac{1}{4}\ln 4=0.6.$$

(3) 直线 $x+y=1.2$ 与曲线 $xy=\frac{1}{4}$ 的交点坐标为（图9）

$$\begin{cases} x_1=0.6+0.1\sqrt{11}=0.932 \\ y_1=0.6-0.1\sqrt{11}=0.268, \end{cases} \qquad \begin{cases} x_2=0.268 \\ y_2=0.932. \end{cases}$$

$\therefore P\{$两数之和小于1.2，两数之积小于$\frac{1}{4}\}$

$$=\frac{\text{阴影区域(Ⅲ)面积}}{\text{正方形面积}}=0.2\times 1+$$

$$+\int_{0.2}^{0.268}(-x+1.2)dx+\int_{0.268}^{0.932}\frac{1}{4x}dx+$$

$$+\int_{0.932}^{1}(-x+1.2)dx$$

$$=0.2+\left(-\frac{1}{2}x^2+1.2x\right)\Big|_{0.2}^{0.268}+$$

$$+\frac{1}{4}\ln x\Big|_{0.268}^{0.932}+\left(-\frac{1}{2}x^2+1.2x\right)\Big|_{0.932}^{1}$$

$$=0.2+0.0657+0.3116+0.0160=0.593$$

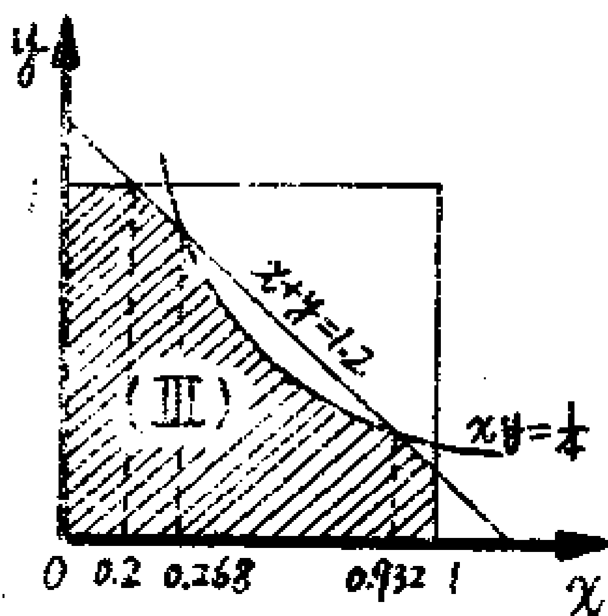


图9

33. 设 A_1, A_2, …, A_n 是随机事件，试用归纳法证明下列公式：

$$P(A_1 \cup A_2 \cup \cdots \cup A_n) = \sum_{i=1}^{n} P(A_i) - \sum_{n \geq j > i \geq 1} P(A_i A_j) +$$

$$\cdots + (-1)^{n-1} P(A_1 A_2 \cdots A_n).$$

证：当 $n=2$ 时，$A_1 \cup A_2 = A_1 \cup (A_2 - A_1 A_2)$，$A_1$ 与 $A_2 - A_1 A_2$ 两者不相容，所以

$$\begin{aligned} P(A_1 \cup A_2) &= P(A_1) + P(A_2 - A_1 A_2) \\ &= P(A_1) + P(A_2) - P(A_1 A_2). \end{aligned}$$

此即当 $n=2$ 时原式成立。

设对 $n-1$ 原式成立，现证对 n 原式也成立。

$$\begin{aligned} P(A_1 \cup \cdots \cup A_{n-1} \cup A_n) &= P\{(A_1 \cup \cdots \cup A_{n-1}) \cup A_n\} \\ &= P(A_1 \cup \cdots \cup A_{n-1}) + P(A_n) \\ &\quad - P\{(A_1 \cup \cdots \cup A_{n-1}) \cap A_n\} \\ &= P(A_1 \cup \cdots \cup A_{n-1}) + P(A_n) \\ &\quad - P\{A_1 A_n \cup A_2 A_n \cup \cdots \cup A_{n-1} A_n\} \end{aligned}$$

对前后两项分别应用归纳假设得

$$\begin{aligned} &P(A_1 \cup \cdots \cup A_{n-1} \cup A_n) \\ &= \{\sum_{n=1}^{n-1} P(A_i) - \sum_{n-1 \geq j > i \geq 1} P(A_i A_j) + \cdots \\ &\quad + (-1)^{n-2} P(A_1 \cdots A_{n-1})\} + P(A_n) \\ &\quad - \{\sum_{i=1}^{n-1} P(A_i A_n) - \sum_{n-1 \geq j > i \geq 1} P(A_i A_n A_j A_n) + \cdots \\ &\quad + (-1)^{n-2} P(A_1 A_n A_2 A_n \cdots A_{n-1} A_n)\} \\ &= \sum_{i=1}^{n} P(A_i) - \sum_{n \geq j > i \geq 1} P(A_i A_j) + \cdots \end{aligned}$$

$+(-1)^{n-1}P(A_1A_2\cdots A_n)$。

至此，原式得证。

34。某班有N个士兵，每人各有一支枪，这些枪外形完全一样，在一次夜间紧急集合中，若每人随机地取走一支枪，问至少有一个人拿到自己的枪的概率。

解：设A_i={第i个战士拿到自己的枪}，$i=1,2,\cdots,N$。A_i之间相容，现用上题公式解。

$$P(A_i)=(N-1)!\times 1/N!=1/N,$$

$$P(A_iA_j)=(N-2)!\times 1\times 1/N!=1/A_N^2\quad(i\neq j),$$

$$\cdots\cdots,P(A_1A_2\cdots A_N)=1/N!\ .$$

由公式得

P{至少有一个战士拿到自己的枪}

$$=P(A_1\cup A_2\cup\cdots\cup A_N)$$

$$=\sum_{i=1}^{N}P(A_i)-\sum_{N\geq j>i\geq 1}P(A_iA_j)+\cdots$$

$$+(-1)^{N-1}P(A_1A_2\cdots A_N)$$

$$=C_N^1\frac{1}{N}-C_N^2\frac{1}{A_N^2}+\cdots+(-1)^{N-1}C_N^N\frac{1}{N!}$$

$$=1-\frac{1}{2!}+\cdots+(-1)^{N-1}\frac{1}{N!}=\sum_{k=1}^{N}\frac{(-1)^{k-1}}{k!}.$$

注：由此可求得，事件"至少有一个战士拿到自己的枪"的对立事件的概率为

P{N个战士中没有一个战士拿到自己的枪}

$$=1-\sum_{k=1}^{N}\frac{(-1)^{k-1}}{k!}=\sum_{k=2}^{N}\frac{(-1)^k}{k!}=\sum_{k=0}^{N}\frac{(-1)^k}{k!}$$

*35．在上题中求恰好有$k(0\leqslant k\leqslant N)$个人拿到自己的枪的概率。

解： 某 k 个指定的战士拿到自己的枪的概率是 $p_1=1/A_N^k$ 。利用上题注（视这里$N-k$为注中N）可得，其余$N-k$个战士都没有拿到自己枪的概率为 $p_2=\sum_{j=0}^{N-k}\frac{(-1)^j}{j!}$。恰有k个战士拿到自己的枪，则这k个战士可以是N个战士中任意的k个战士，从N个战士中选出一组k个战士共有C_N^k种选法，所以事件“恰有 k个战士拿到自己枪”的概率，是事件“某k个指定战士拿到自己的枪，且其余$N-k$个战士没有拿到自己的枪”概率的C_N^k倍，可得

P{恰有k个战士拿到自己的枪}

$$=C_N^k\frac{1}{A_N^k}\sum_{j=0}^{N-k}\frac{(-1)^j}{j!}=\frac{1}{k!}\sum_{j=0}^{N-k}\frac{(-1)^j}{j!}$$

36．考试时共有N张考签，n个学生参加考试$(n\geqslant N)$，被抽过的考签立刻放回，求在考试结束之后，至少有一张考签没有被抽到的概率。

解： 设考签编号为$1,2,\cdots,N$，记事件$A_i=\{$第 i 号考签未被抽到$\}$，则

$$P(A_i)=(N-1)^n/N^n,$$

$$P(A_iA_j)=(N-2)^n/N^n(i\neq j),\cdots\cdots,$$

$$P(A_1A_2\cdots A_N)=(N-N)^n/N^n=0.$$

诸A_i相容，利用第33题公式计算得

P{至少有一张考签未被抽到}

$$=P\{A_1\cup A_2\cup\cdots\cup A_N\}$$

$$= \sum_{i=1}^{N} P(A_i) - \sum_{N \geq j > i \geq 1} P(A_i A_j) + \cdots$$

$$+ (-1)^{N-1} P(A_1 A_2 \cdots A_N)$$

$$= C_N^1 \frac{(N-1)^n}{N^n} - C_N^2 \frac{(N-2)^n}{N^n} + \cdots$$

$$+ (-1)^{N-2} C_N^{N-1} \frac{1}{N^n} + 0$$

$$= \sum_{i=1}^{N-1} (-1)^{i-1} C_N^i \frac{(N-i)^n}{N^n}.$$

37. 甲，乙，丙三人按下面规则进行比赛，第一局由甲，乙参加而丙轮空，由第一局的优胜者与丙进行第二局比赛，而失败者则轮空，比赛用这种方式一直进行到其中一个人连胜两局为止，连胜两局者成为整场比赛的优胜者。若甲，乙，丙胜每局的概率各为$\frac{1}{2}$，问甲，乙，丙成为整场比赛优胜者的概率各是多少？

解： 这种比赛的可能结果，可以用下面方法表示：

aa, acc, acbb, acbaa, acbacc, acbacbb, …

bb, bcc, bcaa, bcabb, bcabcc, bcabcaa, …

其中a表甲胜，b表乙胜，c表丙胜。

在这些结果中，恰巧包含k 个字母的事件发生的概率应为$1/2^k$，如aa发生的概率为$\frac{1}{4}$， acbb 发生的概率为1/16等等。则

$$p(c) = [P(acc) + P(bcc)] + [P(acbacc) + $$
$$+ P(bcabcc)] + \cdots$$
$$= 2 \times \frac{1}{2^3} + 2 \times \frac{1}{2^6} + 2 \times \frac{1}{2^9} + \cdots = \frac{2}{7}.$$

由于甲，乙两人所处的地位是对称的，所以 $p(a) = p(b)$，得

$$p(a) = p(b) = \frac{1}{2}(1 - 2/7) = 5/14.$$

38．父，母，子三人举行比赛，每局总有一人胜一人负（没有和局），每局的优胜者就与未参加此局的人再进行比赛，如果某人首先打胜了两局，则他就是整个比赛的优胜者。由父决定第一局由哪两人参加，其中儿子实力最强，所以父为了使自己得胜的概率达到最大，就决定第一局由它与妻子先比赛。试证父的决策为最优策略（任何一对选手中一人胜对方的概率在整个比赛中是不变的）。

证：设父胜子的概率为 p_1，子胜父的概率为p_2，父胜母，母胜父，母胜子，子胜母的概率分别是 p_3, p_4, p_5, p_6。则诸p_i间有关系：$p_5+p_6=1$，$p_1<p_3$。仿上题，设首局为父对母，比赛的可能结果为：

$$a_3a_1,\ a_3c_2c_6,\ a_3c_2b_5b_4,\ a_3c_2b_5a_3,$$
$$b_4b_5,\ b_4c_6c_2,\ b_4c_6a_1a_3,\ b_4c_6a_1b_4.$$

a 表父胜，但父胜母与父胜子的概率不同，为明确起见，比赛结果中字母附加下标，下标中 i 对应概率p_i，故

$$p_1(a)=P(a_3a_1+P(a_3c_2b_5a_3)+P(b_4c_6a_1a_3)$$
$$=p_3p_1+p_3p_2p_5p_3+p_4p_6p_1p_3.$$

类似地，第一局若父对子，则可得

$$p_2(a)=P(a_1a_3)+P(a_1b_4c_6a_1)+P(c_2b_5a_3a_1)$$
$$=p_1p_3+p_1p_4p_6p_1+p_2p_5p_3p_1.$$

第一局若子对母，则

$$p_3(a)=P(c_6a_1a_3)+P(b_5a_3a_1)$$
$$=p_6p_1p_3+p_5p_3p_1=p_1p_3(p_6+p_5)=p_1p_3.$$

易见 $p_3(a)<p_2(a)$。 由于 $p_1<p_3$，所以 $p_1p_4p_6p_1<p_4p_6p_1p_3$，$p_2p_5p_3p_1<p_3p_2p_5p_3$，因此$p_2(a)<p_1(a)$。

从而 $p_1(a)>p_2(a)>p_3(a)$。

这说明父的决策最优。

39. 给定 $p=P(A)$，$q=P(B)$，$r=P(A\cup B)$，求 $P(A\bar{B})$ 及 $P(\bar{A}\bar{B})$。

解：$P(A\bar{B})=P(A\ \ B)=P(A\cup B-B)$

$$=P(A\cup B)-P(B)=r-q.$$

$$P(\bar{A}\bar{B})=P(\overline{A\cup B})=1-P(A\cup B)=1-r.$$

40. 已知：$P(AB)=P(A)P(B)$，$C\supset AB$，$\bar{C}\supset\bar{A}\bar{B}$，证明：$P(AC)\geqslant P(A)P(C)$。

证：设 $BC=C_1$，

$C(A-B)=C_2$。

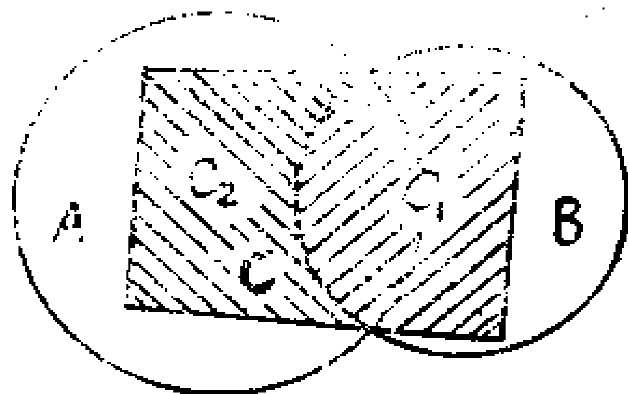


图 10

由 $\bar{C}\supset\bar{A}\bar{B}=\overline{A\cup B}$ 可得

$C\subset A\cup B$，

$\therefore C=C_1\cup C_2$，

$C_1\cap C_2=\phi$ (1)

又 $C\supset AB$

$\therefore AC_1=A(BC)$

$=AB$

再由 $P(B)\geqslant P(C_1)$ 得

$$P(AC_1)=P(AB)=P(A)P(B)\geqslant P(A)P(C_1). \quad (2)$$

由 $C_2\subset A$ 并利用 $P(A)\leqslant 1$ 得

$$P(AC_2)=P(C_2)\geqslant P(A)P(C_2). \quad (3)$$

由(1)，(2)，(3)可得

$$P(AC)=P\{A(C_1\cup C_2)\}=P(AC_1\cup AC_2)$$
$$=P(AC_1)+P(AC_2)$$
$$\geqslant P(A)P(C_1)+P(A)P(C_2)$$
$$=P(A)[P(C_1)+P(C_2)]=P(A)P(C).$$

41.(1)已知 A_1 与 A_2 同时发生则A发生，试证：$P(A) \geqslant P(A_1)+P(A_2)-1$。

(2)若 $A_1A_2A_3 \subset A$，试证 $P(A) \geqslant P(A_1)+P(A_2)+P(A_3)-2$。

证：(1)$A \supset A_1A_2$，由单调性及 $P(A_1 \cup A_2) \leqslant 1$ 得

$$P(A) \geqslant P(A_1A_2) = P(A_1)+P(A_2)-P(A_1 \cup A_2)$$
$$\geqslant P(A_1)+P(A_2)-1.$$

(2)$A \supset A_1A_2A_3$，两次利用(1)的结果得

$$P(A) \geqslant P((A_1A_2)A_3) \geqslant P(A_3)+P(A_1A_2)-1$$
$$\geqslant P(A_3)-1+P(A_1)+P(A_2)-1$$
$$=P(A_1)+P(A_2)+P(A_3)-2.$$

*42. 用概率论想法求 N 阶行列式的展开式中包含主对角线元素的项数。

解：设N阶行列式中元素为 a_{ij}，行列式展开式的每一项为不同行不同列元素的乘积。对于每一项中的各个元素，从第一列中取一个元素有N种取法，当从第一列中取的元素取定后，再从第二列中取一个元素有 N－1 种取法，接着从第三列中取一个元素有N－2种取法，等等。每种取法都是等可能的，共有N！种取法。

设 A_k 表事件{N阶行列式的项含 a_{kk}}，$k=1,2,\cdots,N$，则

$$P(A_k)=(N-1)!\cdot 1/N!=1/N=1/A_N^1,$$

$$P(A_iA_j)=(N-2)!\cdot 1\cdot 1/N!=1/A_N^2 \quad (i \neq j),\cdots\cdots$$

$$P(A_1A_2\cdots A_N)=1/N!\ .$$

至少含一个主对角线元素的项的概率为

$$P\left(\bigcup_{k=1}^{N} A_k\right)=\sum_{k=1}^{N} P(A_k)-\sum_{N \geqslant j > i \geqslant 1} P(A_iA_j)+\cdots+$$
$$+(-1)^{N-1}P(A_1A_2\cdots A_N)$$

$$=C_N^1\frac{1}{A_N^1}-C_N^2\frac{1}{A_N^2}+\cdots+(-1)^{N-1}C_N^N\frac{1}{N!}$$

$$=1-\frac{1}{2!}+\cdots+(-1)^{N-1}\frac{1}{N!}.$$

由此得包含主对角线元素的项数为

$$N!\sum_{k=1}^{N}(-1)^{k-1}\frac{1}{k!}.$$

注：不含主对角线元素项的概率为

$$p_N=1-P(\bigcup_{k=1}^{N}A_k)=\sum_{k=0}^{N}(-1)^k\frac{1}{k!},$$

$$\lim_{N\to\infty}p_N=1/e.$$

43. 利用概率论的想法证明下列恒等式：

$$1+\frac{A-a}{A-1}+\frac{(A-a)(A-a-1)}{(A-1)(A-2)}+\cdots$$

$$+\frac{(A-a)\cdots2\cdot1}{(A-1)\cdots(a+1)a}=\frac{A}{a}.$$

其中A，a都是正整数，且A>a。

证：设袋中有A个球，其中 a 个是白球，不还原随机取出，第k次才首次取得白球的概率为

$$p_k=\frac{A_{A-a}^{k-1}A_a^1}{A_A^k}$$

$$=\frac{a(A-a)(A-a-1)\cdots(A-a-k+2)}{A(A-1)(A-2)\cdots(A-k+1)}$$

$$(k=1,2,\cdots,A-a+1).$$

因为袋中有a 个白球，A－a 个黑球，若一开始总是取到黑

球，直到把黑球取完为止，则至迟到第 $A-a+1$ 次一定会取到白球；也就是说，第一次或第二次…或至迟到第 $A-a+1$次取得白球事件是必然事件，其概率为1。所以

$$1=p_1+p_2+\cdots+p_{A-a+1}$$
$$=\frac{a}{A}+\frac{a(A-a)}{A(A-1)}+\cdots+\frac{a(A-a)\cdots2\cdot1}{A(A-1)\cdots(a+1)a}.$$

等式两边同乘以A/a得

$$1+\frac{A-a}{A-1}+\frac{(A-a)(A-a-1)}{(A-1)(A-2)}+\cdots$$
$$+\frac{(A-a)\cdots2\cdot1}{(A-1)\cdots(a+1)a}=\frac{A}{a}.$$

44. 用某种药物对患有胃溃疡的512个病人进行治疗，结果368人有明显疗效，现有某胃溃疡病人欲服此药，你能对其效果做何估计？

解：有明显疗效的频率为368/512＝71.9%，所以，某胃溃疡病人若服此药，约有71.9%的可能有明显疗效。

45. 求包含事件A，B的最小σ－域。

解：此 σ－域首先包括 Ω,ϕ,A,B 诸元素，然后通过求逆，并，交运算逐步产生新的元素，得共包含如下16个元素：

$$\{\Omega,\phi,A,B,\overline{A},\ \overline{B},AB,A\overline{B},\ \overline{A}B,\ \overline{A}\overline{B},$$
$$AB\cup\overline{A}\overline{B},A\overline{B}\cup\overline{A}B,A\cup B,\ \overline{AB}=\overline{A}\cup\overline{B},$$
$$\overline{A\overline{B}}=\overline{A}\cup B,\ \overline{\overline{A}B}=A\cup\overline{B}\}.$$

证一：现证明它是包含A，B的最小 σ－域。首先它包含A，B；由于所有集均由 A，B 产生，故最小集是有限个，故只需证它为代数，即按如下两条验证集系封闭即可：

1°若 $C\in F$，则 $\overline{C}\in F$；2°若C，$D\in F$，则$C\cup D\in F$。通过验证知确为代数。

证二：由文图知，两个集至多可产生四个部分(图11)，可称之为产生集的最小部分．从这四个部分中任取0，1，2，3，4个求并集，共同构成

$$C_4^0+C_4^1+C_4^2+C_4^3+C_4^4=16$$

个集。故若能找到16个由A，B产生的不同的集，则它们一定是由A，B产生的σ-域，为此只须验证如上16个集两两不同就够了。也可在一开始就根据这16个集的构成法依次构造出来，即得欲求的σ-代数，而不需要再证明。

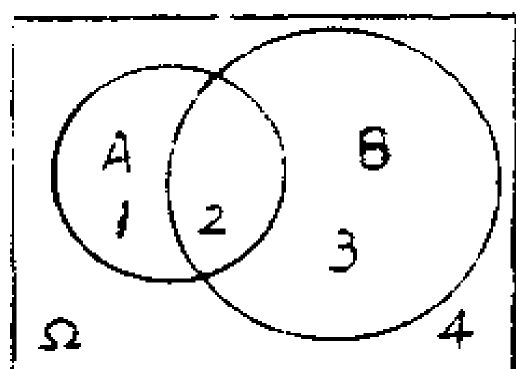


图 11

46。证明Ω的一切子集组成的集类是一个σ-域。

证：记F＝{Ω的一切子集}

(i)Ω是Ω的子集，所以Ω∈F。

(ii)若A∈F，则A是Ω的子集，Ω－A也是Ω的子集，所以$\bar{A}=\Omega-A\in F$。

(iii)$A_i(i=1,2,\cdots)\in F$，当然有$\Omega\supset A_i, i=1,2,\cdots$。任一$\omega\in\bigcup_i A_i$，必有某一$A_i$，使$\omega\in A_i$，所以$\omega\in\Omega$，从而$\Omega\supset\bigcup_i A_i$，即$\bigcup_i A_i$也是Ω的一个子集，故$\bigcup A_i\in F$。

∴F是σ-域。

47。证明：σ-域之交仍为σ-域。

证：设$F_t(t\in T)$是σ-域，记$F=\bigcup_{t\in T}F_t$。

(i)Ω∈每一F_t，所以$\Omega\in\bigcap_{t\in T}F_t$，即Ω∈F。

(ii) $A\in F$，则$A\in$每一F_t，由F_t是σ-域得$\overline{A}\in$每一F_t，所以$\overline{A}\in\bigcup_{t\in T}F_t$，从而$\overline{A}\in F$。

(iii) $A_i(i=1,2,\cdots)\in F$，则诸A_i必属于每一F_t，由于F_t是σ-域，所以$\bigcup_i A_i\in$每一F_t，即$\bigcup_i A_i\in\bigcap_{t\in T}F_t=F$。

∴f是σ-域。

48. 证明：包含一切形如$(-\infty,x)$的区间的最小σ-域是一维波雷尔σ-域。

证：一维波雷尔σ-域$B=m\{[a,b)\}$是由左闭右开区间类产生的σ-域，$\widetilde{B}=M\{(-\infty,x)\}$是由形如$(-\infty,x)$区间类产生的$\sigma$-域。

因为 $[a,b)=(-\infty,b)-(-\infty,a)$

等式左边是$\widetilde{B}$中两个集的差，由此知$\widetilde{B}$包含一切形如$[a,b)$的集，而B是由一切形如$[a,b)$的集类产生的σ-域，所以$\widetilde{B}\supset B$。

又由于$(-\infty,\ x)=\bigcup_{n=1}^{\infty}[x-n,\ x-n+1)$，

等式右边是B中集的可列并，由此知B包含一切形如$(-\infty,x)$的集，与上段同理得$B\supset\widetilde{B}$。

$\therefore\widetilde{B}=B$。

49. 设Q是定义在σ-域上的非负广义实值函数（即可以取有限或无限值的函数），如果它具有可列可加性，并且$Q(\phi)=0$，则称Q为测度。试说明测度概念是算术中计数概念及几何中长度、面积、体积等概念的推广。

解：算术中的计数：以$s(E)$表集合E包含的元素的个数。(1) $s(E)$非负。(2) 对E_i，$i=1,2,\cdots,n$，若任意两个E_i与

$E_j(i\neq j)$都不包含相同的元素，则$s(\bigcup_{i=1}^{n}E_i)=\sum_{i=1}^{n}s(E_i)$，即和集中包含元素的个数等于每个集所包含元素个数之和，集函数$s(E)$具有有限可加性。(3)若 $E=\phi$是空集，它不包含任何元素，则有$s(\phi)=0$。

几何度量中的长度：以 $m(E)$ 表区间的长度。(1)$m(E)$非负。(2)对区间E_i，$i=1,2,\cdots$，若任两个E_i与E_j都不相交，则 $m(\bigcup_{i=1}^{\infty}E_i)=\sum_{i=1}^{\infty}m(E_i)$，$m(E)$ 具有可列可加性。(3)空集ϕ的长度$m(\phi)=0$。

当区间改成区域，长度改成面积或体积时，如上结论也成立。

把算数中计数、几何度量中的共同性质——非负，可列可加性，空集对应值为 0 ——抽象出来，并加以适当地推广，就得到测度的概念。以一维 L-测度为例，$L(a,b)=b-a$，区间(a,b)的L-测度就是区间的长度，利用有限可加性定义由集系$\{(a,b),a,b\in R^1\}$产生的环上的测度，再利用测度延拓就得到了波雷尔域B上的L-测度。

50. 用测度概念解释古典概型、几何概率及概率论公理化结构中关于概率的定义。

解：在概率论公理化结构中，定义在事件域 F 上的集合函数，若满足(i)非负性：$P(A)\geqslant 0$，$A\in F$；(ii)规范性：$P(\Omega)=1$；(iii)可列可加性：若 $A_i\in F$，$i=1,2,\cdots$，$A_iA_j=\phi$，$i\neq j$，则$P(\sum_{i=1}^{\infty}A_i)=\sum_{i=1}^{\infty}P(A_i)$；则称P为 F上的概率。由这三条性质可推得 $P(\phi)=0$。与上题比较可知，定义

在F上的概率P实质上就是定义在F上的规范性测度。

概率的古典定义：$\Omega=\{\omega_1, \cdots, \omega_n\}$；对$A\subset\Omega$定义其概率为 $P(A)=\dfrac{A\text{包含的样本点数}}{\Omega\text{中样本点总数}}$。P具有非负性，有限可加性，$P(\phi)=0$。这里的概率P相当于算术中的计数，所不同的是，P还具有规范性，即$P(\Omega)=1$。这里P实质上是定义在$F=\{\Omega\text{的一切子集}\}$上具有规范性的测度。

几何概率的定义：$G\subset\Omega$，Ω与G都是波雷尔可测集，对G定义其概率为$P(G)=L(G)/L(\Omega)$，其中$L(G)$表示区域G的L-测度。显然P具有非负性。由L-测度具有可列可加性得P也具有可列可加性。另外，$P(\phi)=0$，$P(\Omega)=1$。所以P是定义在$F=\Omega\cap B$上的具有规范性的测度。

第二章　条件概率与统计独立性

本章内容　条件概率，全概率公式，贝叶斯公式，事件独立性，贝努里试验与直线上的随机游动，二项分布与普阿松分布。

1. 把字母M,A,X,A,M分别写在一张卡片上，充分混合后重新排列，问正好得到顺序 MAXAM 的概率是多少？

解：自左往右数，排第 i 个字母的事件记为A_i，则

$P(A_1)=2/5$，$P(A_2|A_1)=2/4$，

$P(A_3|A_2A_1)=\frac{1}{3}$，$P(A_4|A_3A_2A_1)=\frac{1}{2}$，

$P(A_5|A_4A_3A_2A_1)=1$。

所以题中欲求的概率为

$$P(A_1A_2A_3A_4A_5)=P(A_1)P(A_2|A_1)\cdot$$
$$\cdot P(A_3|A_2A_1)P(A_4|A_3A_2A_1)\cdot$$
$$\cdot P(A_5|A_4A_3A_2A_1)=\frac{2}{5}\cdot\frac{2}{4}\cdot\frac{1}{3}\cdot\frac{1}{2}\cdot1=1/30.$$

2. 有三个孩子的家庭中，已知有一个是女孩，求至少有一个男孩的概率。

解：总场合数为$2^3=8$。设A＝{三个孩子中有一女}，B＝{三个孩子中至少有一男}。A的有利场合数为7，AB 的有利场合数为6，所以题中欲求的概率P(B|A)为

$$P(B|A)=\frac{P(AB)}{P(A)}=\frac{6/8}{7/8}=\frac{6}{7}.$$

3. 若M件产品中包含 m 件废品，今在其中任取两件，

求：（1）已知取出的两件中有一件是废品的条件下，另一件也是废品的条件概率；（2）已知两件中有一件不是废品的条件下，另一件是废品的条件概率；（3）取出的两件中至少有一件是废品的概率。

解：(1) M件产品中有m件废品，M－m件正品。设 $A=$｛两件中有一件是废品｝，$B=$｛两件都是废品｝，显然 $A\supset B$，

则 $P(A)=(C_m^1C_{M-m}^1+C_m^2)/C_M^2 \quad P(B)=C_m^2/C_M^2$，

题中欲求的概率为

$$P(B|A)=P(AB)/P(A)=P(B)/P(A)$$

$$=\frac{C_m^2/C_M^2}{(C_m^1C_{M-m}^1+C_m^2)/C_M^2}=\frac{m-1}{2M-m-1}.$$

(2)设 $A=$｛两件中有一件不是废品｝，$B=$｛两件中恰有一件废品｝，显然$B\subset A$，则

$$P(A)=(C_{M-m}^2+C_m^1C_{M-m}^1)/C_M^2,$$

$$P(B)=C_m^1C_{M-m}^1/C_M^2 .$$

题中欲求的概率为

$$P(B|A)=P(AB)/P(A)=P(B)/P(A)$$

$$=\frac{C_m^1C_{M-m}^1/C_M^2}{(C_{M-m}^2+C_m^1C_{M-m}^1)/C_M^2}=\frac{2m}{M+m-1}$$

(3) P｛取出的两件中至少有一件废品｝

$$=(C_m^1C_{M-m}^1+C_m^2)/C_M^2=\frac{m(2M-m-1)}{M(M-1)}.$$

4. 袋中有 a 只黑球，b 只白球，甲乙丙三人依次从袋中

取出一球（取后不放回），试分别求出三人各自取得白球的概率（b⩾3）。

解：A＝{甲取出一球为白球}， B＝{甲取出一球后，乙取出一球为白球}，C＝{甲，乙各取出一球后，丙取出一球为白球}.则

$$P(A)=a/(a+b)$$

甲取出的球可为白球或黑球，利用全概率公式得

$$\begin{aligned}P(B)&=P(A)P(B|A)+P(\bar{A})P(B|\bar{A})\\&=\frac{b}{a+b}\cdot\frac{b-1}{a+b-1}+\frac{a}{a+b}\cdot\frac{b}{a+b-1}\\&=\frac{b}{a+b}.\end{aligned}$$

甲，乙取球的情况共有四种，由全概率公式得

$$\begin{aligned}P(C)&=P(AB)P(C|AB)+P(A\bar{B})P(C|A\bar{B})+\\&\quad+P(AB)P(C|AB)+P(\bar{A}\bar{B})P(C|\overline{AB})\\&=\frac{b(b-1)}{(a+b)(a+b-1)}\cdot\frac{b-2}{a+b-2}+\\&\quad+\frac{ab}{(a+b)(a+b-1)}\cdot\frac{b-1}{a+b-2}+\\&\quad+\frac{ab}{(a+b)(a+b-1)}\cdot\frac{b-1}{a+b-2}+\\&\quad+\frac{a(a-1)}{(a+b)(a+b-1)}\cdot\frac{b}{a+b-2}\\&=\frac{b(a+b-1)(a+b-2)}{(a+b)(a+b-1)(a+b-2)}=\frac{b}{a+b}\end{aligned}$$

5．从{0,1,2,…,9}中随机地取出两个数字，求其和大于10的概率。

解：设B＝{两数之和大于10}， A_i＝{第一个数取到i}，

$i=0,1,\cdots,9$.则$P(A_i)=1/10$，$P(B|A_0)=P(B|A_1)=0$，$P(B|A_i)=(i-1)/9$，$i=2,3,\cdots,5$；$P(B|A_j)=(j-2)9/$，$j=6,7,8,9$.由全概率公式得欲求的概率为

$$P(B)=\sum_{i=0}^{9}P(A_i)P(B|A_i)=\frac{16}{45}=0.356.$$

6. 甲袋中有a只白球，b只黑球，乙袋中有α只白球，β只黑球，某人从甲袋中任取两球投入乙袋，然后在乙袋中任取两球，问最后取出的两球全为白球的概率是多少？

解：设A_1={从甲袋中取出2只白球}，A_2={从甲袋中取出一只白球一只黑球}，A_3={从甲袋中取出2只黑球}，B={从乙袋中取出2只白球}.则由全概率公式得

$$P(B)=P(B|A_1)P(A_1)+P(B|A_2)P(A_2)+$$
$$+P(B|A_3)P(A_3)$$
$$=\frac{C_a^2C_{\alpha+2}^2}{C_{a+b}^2C_{\alpha+\beta+2}^2}+\frac{C_a^1\ C_b^1C_{\alpha+1}^2}{C_{a+b}^2C_{\alpha+\beta+2}^2}+\frac{C_b^2\ C_\alpha^2}{C_{a+b}^2C_{\alpha+\beta+2}^2}$$

7. 设有N个袋子，每个袋子中装有a只黑球，b只白球，从第一袋中取出一球放入第二袋中，然后从第二袋中取出一球放入第三袋中，如此下去，问从最后一个袋子中取出一黑球的概率是多少？

解：A_1={从第一袋中取出一球是黑球}，……，A_i={从第一袋中取一球放入第二袋中，…，再从第$i-1$袋中取一球放入第i袋中，最后从第i袋中取一球是黑球}，$i=1,\cdots,N$.则

$$P(A_1)=a/(a+b),\quad P(\overline{A_1})=b/(a+b).$$

一般设$P(A_k)=a/(a+b)$，则 $P(\overline{A_k})=b/(a+b)$，得

$$P(A_{k+1}) = P(A_{k+1} \mid A_k)P(A_k) + P(A_{k+1} \mid \overline{A}_k)P(\overline{A}_k) = a/(a+b).$$

由数学归纳法得

$$P(A_N) = a/(a+b).$$

8. 飞机有三个不同的部分遭到射击，在第一部分被击中一弹，或第二部分被击中两弹，或第三部分被击中三弹时，飞机才能被击落，其命中率与每一部分的面积成正比，设三个部分的面积的百分比为0.1,0.2,0.7，若已击中两弹，求击落飞机的概率。

解：设 A_1 = {飞机第一部分中两弹}，A_2 = {飞机第二部分中两弹}，A_3 = {飞机第一部分中一弹}，A_4 = {其它情况}，则

$$A_iA_j = \phi (i \neq j), \quad A_1 + A_2 + A_3 + A_4 = \Omega.$$

$$P(A_1) = 0.1 \times 0.1 = 0.01,$$

$$P(A_2) = 0.2 \times 0.2 = 0.004,$$

A_3 = {第一弹中第一部分且第二弹中第二部分，或第一弹中第一部分且第二弹中第三部分，或第一弹中第二部分且第二弹中第一部分，或第一弹中第三分分且第二弹中第一部分}，

$$P(A_3) = 0.1 \times 0.2 + 0.1 \times 0.7 + 0.2 \times 0.1 + 0.7 \times 0.1 = 0.18,$$

$$P(A_4) = 1 - [P(A_1) + P(A_2) + P(A_3)] = 0.77.$$

设B = {飞机被击落}，则

$$P(B \mid A_i) = 1 \ (i = 1,2,3), \quad P(B \mid A_4) = 0.$$

由全概率公式得

$$P(B) = \sum_{i=1}^{4} P(B \mid A_i)P(A_i)$$

$$=0.01+0.04+0.18=0.23.$$

9. 投硬币n回，第一回出正面的概率为c，第二回后每次出现与前一次相同表面的概率为p，求第n回时出正面的概率，并讨论当$n\to\infty$时的情况。

解： 设$A_i=\{$第 i 回出正面$\}$，记$p_i=P(A_i)$，则由题意利用全概率公式得

$$P(A_{i+1})=P(A_{i+1}|A_i)P(A_i)+P(A_{i+1}|\overline{A}_i)P(\overline{A}_i)$$
$$=pp_i+(1-p)(1-p_i)$$
$$=(2p-1)p_i+(1-p).$$

已知$p_1=c$，依次令$i=n-1, n-2, \cdots, 1$可得递推关系式

$$P_n=(2p-1)p_{n-1}+(1-p),$$
$$p_{n-1}=(2p-1)p_{n-2}+(1-p),\cdots,$$
$$p_2=(2p-1)p_1+(1-p)=(2p-1)c+(1-p).$$

解得

$$P_n=(1-p)[1+(2p-1)+(2p-1)^2+\cdots+(2p-1)^{n-2}]+c(2p-1)^{n-1},$$

当$p\neq1$时利用等比数列求和公式得

$$p_n=(1-p)\frac{1-(2p-1)^{n-1}}{1-(2p-1)}+c(2p-1)^{n-1}$$
$$=\frac{1}{2}-\frac{1}{2}(2p-1)^{n-1}+c(2p-1)^{n-1}. \quad (*)$$

1°若$p=1$，则$p_n\equiv C$，$\lim\limits_{n\to\infty}p_n=C$；

2°若$p=0$，则当$n=2k-1$时，$p_n=c$；当$n=2k$时，$p_n=1-c$。

若$c=\frac{1}{2}$，则$p_n\equiv\frac{1}{2}$，$\lim\limits_{n\to\infty}p_n=\frac{1}{2}$；

若$c\neq\frac{1}{2}$，则$c\neq1-c$，$\lim\limits_{n\to\infty}p_n$不存在。

3°若$0<p<1$，则由(*)式可得

$$\lim_{n\to\infty}p_n=\lim_{n\to\infty}\left[\frac{1}{2}-\frac{1}{2}(2p-1)^{n-1}+c(2p-1)^{n-1}\right]=\frac{1}{2}$$

10．甲乙两袋各装一只白球一只黑球，从两袋中各取出一球相交换放入另一袋中，这样进行了若干次。以 p_n，q_n，r_n 分别记在第n次交换后甲袋中将包含两只白球，一只白球一只黑球，两只黑球的概率。试导出p_{n+1}，q_{n+1}，r_{n+1}用p_n，q_n，r_n表出的关系式，利用它们求p_{n+1}，q_{n+1}，r_{n+1}，并讨论当$n\to\infty$时的情况。

解：令A_i，B_i，C_i分别表示第 i 次交换后，甲袋中有两只白球，一白一黑，两黑球的事件，则由全概率公式得

$$\begin{aligned}p_{n+1}&=P(A_{n+1})=P(A_n)P(A_{n+1}\mid A_n)\\&\quad+P(B_n)P(A_{n+1}\mid B_n)+P(C_n)P(A_{n+1}\mid C_n)\\&=0\cdot p_n+\frac{1}{4}q_n+0\cdot r_n=\frac{1}{4}q_n,\\q_{n+1}&=P(B_{n+1})=P(A_n)P(B_{n+1}\mid A_n)\\&\quad+P(B_n)P(B_{n+1}\mid B_n)+P(C_n)P(B_{n+1}\mid C_n)\\&=1\cdot p_n+\frac{1}{2}q_n+1\cdot r_n=p_n+\frac{1}{2}q_n+r_n,\\r_{n+1}&=P(C_{n+1})=P(A_n)P(C_{n+1}\mid A_n)+\\&\quad+P(B_n)P(C_{n+1}\mid B_n)+P(C_n)P(C_{n+1}\mid C_n)\\&=0\cdot p_n+\frac{1}{4}q_n+0\cdot r_n=\frac{1}{4}q_n.\end{aligned}$$

这里有$p_{n+1}=r_{n+1}$，又$p_{n+1}+q_{n+1}+r_{n+1}=1$，所以$q_{n+1}=1-2p_{n+1}$。同理有 $q_n=1-2p_n$，再由 $p_{n+1}=\frac{1}{4}q_n$ 得 $p_{n+1}=\frac{1}{4}(1-2p_n)$。所以可得递推关系式为

$$\begin{cases}r_{n+1}=p_{n+1}=\frac{1}{4}(1-2p_n),\\q_{n+1}=1-2p_{n+1}\end{cases}$$

初始条件是甲袋一白一黑，乙袋一白一黑，即 $p_0=r_0=0$，$q_0=1$。由递推关系式得

$$r_{n+1}=p_{n+1}=\tfrac{1}{4}(1-2p_n)=\tfrac{1}{4}-\tfrac{1}{2}p_n$$
$$=\tfrac{1}{4}-\tfrac{1}{2}(\tfrac{1}{4}-\tfrac{1}{2}p_{n-1})=\tfrac{1}{4}-\tfrac{1}{8}+\tfrac{1}{4}p_{n-1}=\cdots$$
$$=\frac{1}{2^2}-\frac{1}{2^3}+\cdots+(-1)^{n+2}/2^{n+2}+$$
$$+(-1)^{n+1}p_0/2^{n+1}$$
$$=\frac{\tfrac{1}{4}[1-(-\tfrac{1}{2})^{n+1}]}{1-(-\tfrac{1}{2})}$$
$$=\tfrac{1}{6}[1-(-1)^{n+1}\cdot(\tfrac{1}{2})^{n+1}]$$
$$=\tfrac{1}{6}+(-1)^n\cdot\tfrac{1}{3}\cdot(\tfrac{1}{2})^{n+2},$$
$$q_{n+1}=1-2p_{n+1}=2/3+(-1)^{n+1}\cdot\tfrac{1}{3}\cdot(\tfrac{1}{2})^{n+1}.$$
$$\lim_{n\to\infty}p_n=\lim_{n\to\infty}r_n=\tfrac{1}{6},\quad \lim_{n\to\infty}q_n=2/3.$$

11．设一个家庭中有n个小孩的概率为

$$p_n=\begin{cases}\alpha p^n, & n\geqslant 1,\\ 1-\dfrac{\alpha p}{1-p}, & n=0,\end{cases}$$

这里$0<p<1$，$0<\alpha<(1-p)/p$。若认为生一个小孩为男孩或女孩是等可能的，求证一个家庭有$k(k\geqslant 1)$个男孩的概率为$2\alpha p^k/(2-p)^{k+1}$。

解： 设A_n={家庭中有n个孩子}，$n=0,1,2,\cdots$，B={家庭中有k个男孩}。注意到生男孩与生女孩是等可能的，由二项分布$(p=\tfrac{1}{2})$得

$$P(B|A_n)=C_n^k(\tfrac{1}{2})^k(\tfrac{1}{2})^{n-k}=C_n^k(\tfrac{1}{2})^n.$$

由全概率公式得

$$P(B)=\sum_{n=k}^{\infty}P(A_n)P(B\mid A_n)=\sum_{n=k}^{\infty}\alpha p^n C_n^k(\tfrac{1}{2})^n$$
$$=\alpha\sum_{i=0}^{\infty}C_{k+i}^i\left(\frac{p}{2}\right)^{k+i}\ (\text{其中}i=n-k)$$

$$=\alpha\left(\frac{p}{2}\right)^k \sum_{i=0}^{\infty} C_{k+i}^{i}\left(\frac{p}{2}\right)^i$$

$$=\alpha\left(\frac{p}{2}\right)^k\left(1-\frac{p}{2}\right)^{-k-1}=2\alpha p^k/(2-p)^{k+1}.$$

12. 在上题假设下：(1)已知家庭中至少有一个男孩，求此家庭至少有两个男孩的概率；(2)已知家庭中没有女孩，求正好有一个男孩的概率。

解：(1)设A = {至少有一男孩}，B = {至少有2个男孩}，$A\supset B$，$AB=B$，由$0<p/(2-p)<1$得

$$P(A)=\sum_{k=1}^{\infty}\frac{2\alpha p^k}{(2-p)^{k+1}}=\frac{2\alpha}{2-p}\cdot\frac{p/(2-p)}{1-p/(2-p)}$$

$$=\frac{\alpha p}{(2-p)(1-p)},$$

$$P(B)=\sum_{k=2}^{\infty}\frac{2\alpha p^k}{(2-p)^{k+1}}=\frac{2\alpha}{2-p}\cdot\frac{p^2/(2-p)^2}{1-p/(2-p)}$$

$$=\frac{\alpha p^2}{(2-p)^2(1-p)},$$

$$P(B\mid A)=\frac{P(AB)}{P(A)}=\frac{P(B)}{P(A)}=\frac{p}{2-p}.$$

(2)C = {家中无女孩} = {家中无小孩，或家中有 n 个小孩且都是男孩，n是任意正整数}，则

$$P(C)=1-\frac{\alpha p}{1-p}+\sum_{n=1}^{\infty}\alpha p^n\left(\frac{1}{2}\right)^n$$

$$=1-\frac{\alpha p}{1-p}+\frac{\alpha p/2}{1-p/2}=1-\frac{\alpha p}{1-p}+\frac{\alpha p}{2-p}$$

$$=\frac{2-3p-\alpha p+p^2}{(1-p)(2-p)}$$

A_1 = {家中正好有一个男孩} = {家中只有一个小孩且

是男孩}，则

$$P(A_1)=\alpha p\cdot\frac{1}{2}=\frac{1}{2}\alpha p,\ 且A_1\subset C,$$

所以在家中没有女孩的条件下，正好有一个男孩的条件概率为

$$P(A_1|C)=\frac{P(A_1C)}{P(C)}=\frac{P(A_1)}{P(C)}$$

$$=\frac{1}{2}\alpha p\Big/\frac{2-3p-\alpha p+p^2}{(1-p)(2-p)}$$

$$=\frac{\alpha p(1-p)(2-p)}{2(2-3p-\alpha p+p^2)}$$

13. 已知产品中96%是合格品，现有一种简化的检查方法，它把真正的合格品确认为合格品的概率为0.98，而误认废品为合格品的概率为0.05，求在简化方法检查下，合格品的一个产品确实是合格品的概率。

解：设 A={产品确为合格品}，B={检查后判为合格品}.已知$P(B|A)=0.98$，$P(B|\bar{A})=0.05$，$P(A)=0.96$，求$P(A|B)$.由贝叶斯公式得

$$P(A|B)=\frac{P(AB)}{P(B)}$$

$$=\frac{P(A)P(B|A)}{P(A)P(B|A)+P(\bar{A})P(B|\bar{A})}$$

$$=\frac{0.96\times0.98}{0.96\times0.98+0.04\times0.05}=\frac{0.9408}{0.9428}$$

$$=0.9979.$$

14. 炮战中，在距目标250米，200米，150米处射击的概率分别为0.1，0.7，0.2，而在各该处射击时命中目标的概率分别为0.05，0.1，0.2，现在已知目标被击毁，求击毁目标的炮弹是由250米处射出的概率。

原书缺页

原书缺页

$$P(\bar{A}\bar{C})=P(\bar{A})P(\bar{C}),$$
$$P(\bar{B}\bar{C})=P(\bar{B})P(\bar{C}).$$

又有

$$\begin{aligned}P(\bar{A}\bar{B}\bar{C}) &= P(\overline{A\cup B\cup C})=1-P(A\cup B\cup C)\\ &=1-[P(A)+P(B)+P(C)-P(AB)-\\ &\quad -P(AC)-P(BC)+P(ABC)]\\ &=1-P(A)-P(B)-P(C)+P(A)P(B)+\\ &\quad +P(A)P(C)+P(B)P(C)-P(A)P(B)P(C)\\ &=(1-P(A))(1-P(B))(1-P(C))\\ &=P(\bar{A})P(\bar{B})P(\bar{C}),\end{aligned}$$

所以$\bar{A}$，$\bar{B}$，$\bar{C}$相互独立。

18．证明：事件 $A_1,A_2\cdots,A_n$ 相互独立的充要条件是下列2^n个等式成立：

$$P(\hat{A}_1\ \hat{A}_2\ \cdots\ \hat{A}_n)=P(\hat{A}_1)P(\hat{A}_2)\cdots P(\hat{A}_n),$$

其中$\hat{A}_i$取 A_i 或 $\bar{A}_i$。

证：必要性．设事件$A_1,A_2,\cdots,A_n$相互独立，用归纳法证．不失一般性，假设总是前连续 m 个集$\hat{A}_i$取$\bar{A}_i$ 的形式．当$m=1$时，

$$\begin{aligned}P(\bar{A}_1A_2\cdots A_n) &= P(A_2\cdots A_n)-P(A_1\cdots A_n)\\ &=P(A_2)\cdots P(A_n)-P(A_1)\cdots P(A_n)\\ &=P(\bar{A}_1)P(A_2)\cdots P(A_n).\end{aligned}$$

设当$m=k$时有

$$P(\bar{A}_1\cdots\bar{A}_kA_{k+1}\cdots A_n)=$$
$$=P(\bar{A}_1)\cdots P(\bar{A}_k)P(A_{k+1}\cdots A_n),$$

则当$m=k+1$时

$$P(\bar{A}_1\cdots\bar{A}_{k+1}A_{k+2}\cdots A_n)=$$

$$=P(\bar{A}_1\cdots\bar{A}_kA_{k+2}\cdots A_n)$$

$$-P(\bar{A}_1\cdots\bar{A}_kA_{k+1}\cdots A_n)$$

$$=P(\bar{A}_1)\cdots P(\bar{A}_k)P(A_{k+2})\cdots P(A_n)-$$

$$-P(\bar{A}_1)\cdots P(\bar{A}_k)P(A_{k+1})\cdots P(A_n)$$

$$=P(\bar{A}_1)\cdots P(\bar{A}_k)(1-P(A_{k+1}))P(A_{k+2})\cdots P(A_n)$$

$$=P(\bar{A}_1)\cdots P(\bar{A}_k)P(\bar{A}_{k+1})P(A_{k+2})\cdots P(A_n)$$

从而有下列2^n个等式成立：

$$P(\hat{A}_1\hat{A}_2\cdots\hat{A}_n)=P(\hat{A}_1)P(\hat{A}_2)\cdots P(\hat{A}_n),$$

其中$\hat{A}_i$取A_i或$\bar{A}_i$。

充分性。设题中条件成立，则

$$P(A_1\cdots A_n)=P(A_1)\cdots P(A_n),\qquad(1)$$

$$P(A_1\cdots A_{n-1}\bar{A}_n)=P(A_1)\cdots P(A_{n-1})P(\bar{A}_n).\qquad(2)$$

∵ $A_1\cdots A_{n-1}A_n\cap A_1\cdots A_{n-1}\bar{A}_n=\phi$,

∴ $P(A_1\cdots A_{n-1})=P(A_1\cdots A_{n-1}A_n\cup A_1\cdots A_{n-1}\bar{A}_n)$.

(1)+(2)得

$$P(A_1\cdots A_{n-1})=P(A_1)\cdots P(A_{n-1}).\qquad(3)$$

同理有

$$P(A_1\cdots A_{n-2}\bar{A}_{n-1}A_n)=$$

$$=P(A_1)\cdots P(A_{n-2})P(\bar{A}_{n-1})P(A_n),$$

$$P(A_1\cdots A_{n-2}\bar{A}_{n-1}\bar{A}_n)=$$

$$=P(A_1)\cdots P(A_{n-2})P(\bar{A}_{n-1})P(\bar{A}_n)$$

两式相加得

$$P(A_1\cdots A_{n-2}\bar{A}_{n-1})=P(A_1)\cdots P(A_{n-2})P(\bar{A}_{n-1}).\qquad(4)$$

(3)+(4)得

$$P(A_1\cdots A_{n-2})=P(A_1)P(A_2)\cdots P(A_{n-2}).$$

原
书
缺
页

原书缺页

确切意思。

解： 以 A_k 表事件"A于第 k 次试验中出现"， $P(A_k)=\varepsilon$，由试验的独立性得，前 n 次试验中A都不出现的概率为

$$P(\bar{A}_1\bar{A}_2\cdots\bar{A}_n)=P(\bar{A}_1)P(\bar{A}_2)\cdots P(\bar{A}_n)$$
$$=(1-\varepsilon)^n.$$

于是前 n 次试验中，A至少发生一次的概率为

$$1-P(\bar{A}_1\bar{A}_2\cdots\bar{A}_n)=1-(1-\varepsilon)^n\to 1\ (n\to\infty).$$

这说明当重复试验的次数无限增加时，小概率事件A至少发生一次的概率可以无限地向 1 靠近，从而可看成是必然要发生的。

24. 在第一台车床上制造一级品零件的概率等于0.7，而在第二台车床上制造此种零件的概率等于0.8，第一台车床制造了两个零件，第二台制造了三个零件，求所有零件均为一级品的概率。

解： 我们认为，各车床或同一车床制造的各个零件的好坏是相互独立的，由此可得

$$P\{\text{所有零件均为一级品}\}=0.8^3\times 0.7^2=0.2509.$$

25. 实验室器皿中产生甲类细菌与乙类细菌的机会是相同的，若某次发现产生了2n个细菌，求(1)至少有一个甲类细菌的概率；(2)甲，乙两类细菌各占其半的概率。

解： 利用$p=\frac{1}{2}$的二项分布可得

$$P\{\text{至少有一个甲类细菌}\}=1-P\{2n\text{个全是乙类细菌}\}$$
$$=1-C_{2n}^0\left(\frac{1}{2}\right)^0\left(\frac{1}{2}\right)^{2n}=1-2^{-2n}.$$

$$P\{\text{甲，乙两类细菌各占一半}\}=$$
$$=C_{2n}^n\left(\frac{1}{2}\right)^n\left(\frac{1}{2}\right)^n=C_{2n}^n\left(\frac{1}{2}\right)^{2n}.$$

26. 掷硬币出现正面的概率为 p，掷了 n 次，求下列概率：(1)至少出现一次正面；(2)至少出现两次正面。

解： 利用二项分布得

$$P\{至少出现一次正面\}=1-P\{n次全部出现反面\}$$
$$=1-(1-p)^n.$$

$$P\{至少出现两次正面\}=$$
$$=1-(1-p)^n-C_n^1p(1-p)^{n-1}$$
$$=1-(1-p)^n-np(1-p)^{n-1}.$$

27. 甲，乙，丙三人进行某项比赛，设三个胜每局的概率相等，比赛规定先胜三局者为整场比赛的优胜者，若甲胜了第一，三局，乙胜了第二局，问丙成为整场比赛优胜者的概率是多少？

解： (i)设A，B，C 分别表示每局比赛中甲，乙，丙获胜的事件，这是一个$P(A)=P(B)=P(C)=\frac{1}{3}$的多项分布。欲丙成为整场比赛的优胜者，则需在未来的三次中，丙获胜三次；或在前三次中，丙胜两次乙胜一次，而第四次为丙获胜.故本题欲求的概率为

$$p=\frac{3!}{3!0!0!}\left(\frac{1}{3}\right)^3\left(\frac{1}{3}\right)^0\left(\frac{1}{3}\right)^0+$$
$$+\frac{3!}{2!1!0!}\left(\frac{1}{3}\right)^2\left(\frac{1}{3}\right)\left(\frac{1}{3}\right)^0\cdot\frac{1}{3}=\frac{1}{27}+\frac{1}{27}=\frac{2}{27}.$$

28. 甲，乙均有n 个硬币，全部掷完后分别计算掷出的正面数，试求两人掷出的正面数相等的概率。

解： 利用两个$p=\frac{1}{2}$的二项分布，得欲求的概率为

$$p=\sum_{i=0}^{n}P\{甲掷出\ i\ 次正面，乙掷出\ i\ 次正面\}$$

原书缺页

原书缺页

$$+\frac{1}{N}\left(\frac{N-2}{N}\right)^{n-1}+\left(\frac{N-2}{N}\right)^{n}$$

$$=\frac{1}{N}\left[1-\left(\frac{N-2}{N}\right)^{n}\right]\Big/\left(1-\frac{N-2}{N}\right)+\left(\frac{N-2}{N}\right)^{n}$$

$$=\frac{1}{2}+\frac{1}{2}\left(\frac{N-2}{N}\right)^{n}.$$

若$N=1$，则当$n=2k+1$时$p_n=0$，当$n=2k$时$p_n=1$。

若$N=2$，则对任何 n 有 $p_n=\frac{1}{2}$。

若$N>2$，则$\lim\limits_{n\to\infty}p_n=\frac{1}{2}$（N越大，收敛速度越慢）。

32。一个工厂出产的产品中废品率为0.005，任意取来1000件，试计算下面概率：(1)其中至少有两件废品；(2)其中不超过 5 件废品；(3)能以90%的概率希望废品件数不超过多少？

解：利用普阿松逼近定理，$\lambda=1000\times0.005=5$，查表计算得

P{至少有两件废品}=

$$=\sum_{i=2}^{1000}C_{1000}^{i}(0.005)^{i}(0.995)^{1000-i}$$

$$\approx1-e^{-5}-5e^{-5}=0.9596.$$

P{不超过5件废品}=

$$=\sum_{i=0}^{5}C_{1000}^{i}(0.005)^{i}(0.995)^{1000-i}\approx\sum_{i=0}^{5}\frac{5^{i}}{i!}e^{-5}=0.6160.$$

设以90%的概率希望废品件数不超过k，则

$$\sum_{i=0}^{k}C_{1000}^{i}(0.005)^{i}(0.995)^{1000-i}\approx\sum_{i=0}^{k}\frac{5^{i}}{i!}e^{-5}=0.90,$$

解得$k=8$。

33. 某交往式计算机有20个终端，这些终端被各单位独立操作，使用率各为0.7，求有10个或更多个终端同时操作的概率。

解： P{有10个或更多个终端同时操作}=

=P{有10个或不足10个终端不在操作}

$$=\sum_{i=0}^{10} C_{20}^{i}(0.3)^{i}(0.7)^{20-i}=0.9829.$$

最后一步是利用教本P88表1中$p_2=0.3$一栏求值的。

34. 设每次射击打中目标的概率等于0.001，如果射击5000次，试求打中两弹或两弹以上的概率。

解： 利用普阿松逼近定理计算，$\lambda=5000\times0.001=5$，则打中两弹或两弹以上的概率为

$$p=1-(0.999)^{5000}-5000(0.999)^{4999}\times0.001$$
$$\approx 1-e^{-5}-5e^{-5}=0.9596.$$

35. 某个厂有7个顾问，假定每个顾问贡献正确意见的百分比为0.6，现为某事可行与否而个别征求顾问意见，并按多数人的意见作出决策，求作出正确决策的概率。

解： 设A表事件"某事实际上是可行的"，$\overline{A}$表事件"某事实际上是不可行的"，B表"多数人说可行"$\overline{B}$表"多数人说不可行"，利用二项分布得

$$P(B\mid A)=P(\overline{B}\mid\overline{A})=\sum_{i=4}^{7} C_7^{i}(0.6)^{i}(0.4)^{7-i}=0.7102$$

所以作出正确决策的概率为

$$p=P(A)P(B\mid A)+P(\overline{A})P(\overline{B}\mid\overline{A})$$
$$=P(B\mid A)[P(A)+P(\overline{A})]=P(B\mid A)=0.7102.$$

原
书
缺
页

原书缺页

$$p=\sum_{i=0}^{k-1}C_{100+k}^{i}(0.015)^{i}(0.985)^{100+k-i}\geqslant 0.80.$$

由于k值不大，有

$$(100+k)0.015\approx 100\times 0.015=1.5,$$

利用普阿松逼近定理计算，$\lambda=1.5$，上式可以写成

$$p=\sum_{i=0}^{k-1}\frac{(1.5)^{i}}{i!}e^{-1.5}\geqslant 0.80.$$

查表得当 $k-1=2$ 时，$p=0.808847$；当 $k-1=1$时，$p=0.557825$.取 $k-1=2$，$k=3$. 所以一盒应装103只，才能保证每盒中有100只以上好钉的概率不小于80%.

41. 某疫苗中所含细菌数服从普阿松分布，每 1 毫升中平均含有一个细菌，把这种疫苗放入 5 只试管中，每试管放 2 毫升，试求：(1) 5 只试管中都有细菌的概率；(2)至少有 3 只试管中有细菌的概率.

解：每一毫升平均含一个细菌，每 2 毫升含 2 个，所以每只试管中含有细菌数服从$\lambda=2$的普阿公分布.由此可得

P{5个试管中都有细菌} $=(1-e^{-2})^{5}=0.4833$；

P{至少有三个试管中有细菌} =

$$=\sum_{i=3}^{5}C_{5}^{i}\ (1-e^{-2})^{i}(e^{-2})^{5-i}=0.9800.$$

计算时利用了$p=1-e^{-2}$的二项分布.

42. 通过某交叉路口的汽车流可看作普阿松过程，若在一分钟内没有车的概率为0.2，求在2分钟内有多于一车的概率.

解：设一分钟内通过某交叉路口的汽车数服从λ_1的普阿

松分布．则

$$P\{1\text{分钟内无车}\}=e^{-\lambda_1}=0.2,\ \lambda_1=-\ln 0.2=1.61.$$

由此得，2分钟内通过的汽车数服从 $\lambda=\lambda_1\times 2=3.22$的普阿松分布，从而2分钟内多于一车的概率为

$$p=1-e^{-3.22}-3.22\times e^{-3.22}=0.831.$$

43．若每蚕产 n个卵的概率服从普阿松分布，参数为λ，而每个卵变为成虫的概率为 p，且各卵是否变为成虫彼此间没有关系，求每蚕养出 k 只小蚕的概率。

解： 若蚕产 i 个卵，则这 i 个卵 变为成虫数服从概率为 p，$n=i$ 的二项分布，所以

$$P\{\text{蚕养出 k 只小蚕}\}=$$

$$=\sum_{i=k}^{\infty}\frac{\lambda^i}{i!}e^{-\lambda}\ C_i^k\ p^k(1-p)^{i-k}\ (\text{令}m=i-k)$$

$$=\frac{p^k e^{-\lambda}}{k!}\sum_{m=0}^{\infty}\frac{\lambda^{k+m}}{m!}(1-p)^m=\frac{1}{k!}(\lambda p)^k e^{-\lambda p}.$$

44．若已知 $t=0$ 时，某分子与另一分子碰撞，又知对任何 $t\geqslant 0$和 $\Delta t>0$，若不管该分子在时刻 t 以前是否遭受碰撞，在$(t,t+\Delta t)$中遭到碰撞的概率等于$\lambda\Delta t+o(\Delta t)$，试求该分子在时刻 τ 还没有再受到碰撞的概率。

解： 设s＝{该分子在时刻 s 还没有再受到碰撞}，则

$$1-P(\Delta\tau)=\lambda(\Delta\tau)+o(\Delta\tau),$$

$$P(\tau+\Delta\tau)=P(\tau)P(\Delta\tau)=P(\tau)(1-\lambda\Delta\tau-o(\Delta\tau)),$$

$$\frac{P(\tau+\Delta\tau)-P(\tau)}{\Delta\tau}=-\lambda p(\tau)-\frac{P(\tau)o(\Delta\tau)}{\Delta\tau},$$

令$\Delta\tau\to 0$得

$$\frac{dP(\tau)}{d\tau}=-\lambda P(\tau),\quad \frac{P'(\tau)}{P(\tau)}=-\lambda,$$

积分得

$$P(\tau)=ce^{-\lambda\tau}.$$

当$\tau\to 0$时，$P(\tau)\to 1$，所以$c=1$，从而

$$P(\tau)=e^{-\lambda\tau}.$$

*45. 利用概率论的想法证明下面恒等式：

$$\sum_{k=0}^{M}\binom{N+k}{k}\frac{1}{2^k}=2^N.$$

证： 可利用巴纳赫氏问题证明.某数学家带着两盒火柴，每次用时他在两盒中任意抓一盒，从中取出一根，因此连续地抽取构成了一串 $p=\frac{1}{2}$ 的贝努里试验.假定最初每盒火柴恰巧包含 N根，我们考虑：数学家第一次发现空盒子的时刻.在那一时刻，另一盒火柴可能还有 r 为 $0,1,\cdots,N$ 根火柴.设从第一盒中选取为"成功"."当发现第一盒火柴空时，第二盒中尚有 r 根火柴"这一事件，等价于"恰有$N-r$ 次失败发生在第 $N+1$ 次成功之前"，这个事件的概率为 $f(2N-r+1;N+1,\frac{1}{2})$（见巴斯卡分布）.考虑到两盒火柴所处的地位相同，可得事件"发现一盒空，另一盒中尚有 r 根火柴"（记为A_r）的概率为

$$2f(2N-r+1;N+1,\tfrac{1}{2})=$$

$$=2\binom{2N-r}{N}2^{-2N+r+1}=\binom{2N-r}{N}2^{-2N+r}.$$

r 取 0 到N的诸事件A_r之和显然是必然事件，由此可得

$$\sum_{r=0}^{N}\binom{2N-r}{N}2^{-2N+r}=1,$$

两边同乘以2^N并利用组合性质变形得

$$\sum_{r=0}^{N}\binom{2N-r}{N-r}2^{-(N-r)}=2^N,$$

令$N-r=k$，并注意到对应 r 从 0 变到N，而 k 是从N变到0，即得要证的等式

$$\sum_{k=0}^{N}\binom{N+k}{k}2^{-k}=2^N.$$

*46．通过构造适当的概率模型证明：从正整数中随机地选取两数，此两数互素的概率等于$6/\pi^2$。

证： 任何一个非 1 的自然数，皆可唯一地(不计次序时)分解为素数的乘积，要证两数互素，只需验证这两数没有公共素因子就行了。为此，把素数排列为 $p_1<p_2<\cdots$，对任何 t，N（自然数）定义事件

$A_{N,t}=\{$在$1,2\cdots,N$中独立地取两整数ξ,η，ξ 与η 不含公因子$p_1,p_2,\cdots,p_t\}$。把所要求的"事件"的概率定义为

$$\lim_{t\to\infty}(\lim_{N\to\infty}P(A_{N,t})).$$

为计算$P(A_{N,t})$，定义

$m_{i_1\cdots i_k}=P\{$自$1,2,\cdots,N$中独立地取两整数ξ,η，它们有公因子$p_{i_1},p_{i_2},\cdots,p_{i_k}\}$.

则由事件容许相容的和的概率公式得

$$P(A_{N,t})=1-\sum_{i=1}^{t}m_i+\sum_{t\geq j>i\geq 1}m_{ij}-\cdots+(-1)^t m_{12\ldots t}. \tag{1}$$

显然有 $m_{i_1\cdots i_k}=\left(\frac{1}{N}\left[N/p_{i_1}\cdots p_{i_k}\right]\right)^2$，

$$\left(\frac{1}{p_{i_1}\cdots p_{i_k}}-\frac{1}{N}\right)^2\leqslant m_{i_1\cdots i_k}\leqslant\left(\frac{1}{p_{i_1}\cdots p_{i_k}}\right)^2,$$

因而

$$\sum_{1\leqslant i_1<\cdots<i_k\leqslant t}\frac{1}{p_{i_1}^2\cdots p_{i_k}^2}-\frac{2C_t^k}{N}$$

$$\leqslant\sum_{1\leqslant i_1<\cdots<i_k\leqslant t}m_{i_1\cdots i_k}$$

$$\leqslant\sum_{1\leqslant i_1<\cdots<i_k\leqslant t}\frac{1}{p_{i_1}^2\cdots p_{i_k}^2}. \qquad (2)$$

(2)式左端的 $2C_t^k/N$的来由是，

$$\left(\frac{1}{p_{i_1}\cdots p_{i_k}}-\frac{1}{N}\right)^2\geqslant\frac{1}{p_{i_1}^2\cdots p_{i_k}^2}-\frac{1}{N}\cdot\frac{2}{p_{i_1}\cdots p_{i_k}}$$

$$\geqslant\frac{1}{p_{i_1}^2\cdots p_{i_k}^2}-\frac{2}{N},$$

而和式中一共有C_t^k项。由(1)，(2)得

$$1-\sum_{i=1}^{t}\frac{1}{p_i^2}+\sum_{t\geqslant j>i\geqslant 1}\frac{1}{p_i^2\ p_j^2}-\cdots-\frac{2}{N}\sum_{k=1}^{t}C_t^k$$

$$\leqslant P(A_{N,t})\leqslant$$

$$\leqslant 1-\sum_{i=1}^{t}\frac{1}{p_i^2}+\sum_{t\geqslant j>i\geqslant 1}\frac{1}{p_i^2\ p_j^2}-\cdots+\frac{2}{N}\sum_{k=1}^{t}C_t^k. \quad (3)$$

在(3)中令$N\to\infty$得

$$\lim_{N\to\infty}P(A_{N,t})=1-\sum_{i=1}^{t}\frac{1}{p_i^2}+\sum_{t\geq j>i\geq 1}\frac{1}{p_i^2\ p_j^2}-\cdots$$

$$(-1)^t\frac{1}{p_1^2\cdots p_t^2}=\prod_{i=1}^{t}\left(1-\frac{1}{p_i^2}\right),$$

再令$t\to\infty$，并利用黎曼函数$\zeta(2)=6/\pi^2$（参看华罗庚著“数论导引”p236，225）得，欲求的概率为

$$\lim_{t\to\infty}[\lim_{N\to\infty}P(A_{N,t})]=\prod_{i=1}^{\infty}\left(1-\frac{1}{p_i^2}\right)=\zeta(2)=\frac{6}{\pi^2}$$

47．某车间宣称自己产品的合格率超过99%，检验人员从该车间的10000件产品中抽查了100件，发现有两件次品，能否据此断定该车间谎报合格率？

解：假设产品合格率 $p\geqslant 0.99$，不妨设 $p=0.99$。现从10000件中抽100件，可视为放回抽样．而100件产品中次品件数服从二项分布，利用普阿松逼近定理得，次品件数不小于两件的概率为

$$p=1-(0.99)^{100}-100\times 0.01\times 0.99^{99}$$
$$\approx 1-e^{-1}-e^{-1}=0.2642.$$

此非小概率事件，所以不能据此断定该车间慌报合格率．(注意，这并不代表可据此断定，该车间没有慌报合格率．)

第三章　随机变量与分布函数

本章内容　随机变量及其分布，随机向量，随机变量的独立性，随机变量的函数及其分布。

1．直线上有一质点，每经一个单位时间，它分别以概率p及1－p向右或向左移动一格，若该质点在时刻0从原点出发，而且每次移动是相互独立的，试用随机变量来描述这质点的运动（以S_n表示时刻n时质点的位置）。

解：令ξ_n表在n次移动中向右移动的次数，则ξ_n服从二项分布，

$$P\{\xi_n=k\}=C_n^k\ p^k(1-p)^{n-k},\ k=0,1,\cdots,n$$

以S_n表时刻n时质点的位置，则

$$S_n=\xi_n-(n-\xi_n)=2\xi_n-n.$$

ξ_n的分布列为

$$\begin{pmatrix} 0 & 1 & 2 & \cdots & n \\ (1-p)^n & C_n^1p(1-p)^{n-1} & C_n^2p^2(1-p)^{n-2} & \cdots & p^n \end{pmatrix}.$$

S_n的分布列为

$$\begin{pmatrix} -n & -n+2 & -n+4 & \cdots & n \\ (1-p)^n & C_n^1p(1-p)^{n-1} & C_n^2p^2(1-p)^{n-2} & \cdots & p^n \end{pmatrix}.$$

2．设ξ为贝努里试验中第一个游程（连续的成功或失败）

的长，试求 ξ 的概率分布。

解：$P\{\xi=1\}=P\{失成\}+P\{成失\}=pq+qp$，

$P\{\xi=2\}=P\{失失成\}+P\{成成失\}=ppq+qqp$
$=p^2q+q^2p$，…

所以 ξ 的概率分布为

$P\{\xi=k\}=p^kq+q^kp$，$k=1,2,\cdots$。

3．c应取何值才能使下列函数成为概率分布：(1) $f(k)=\frac{c}{N}$，$k=1,2,\cdots,N$；(2) $f(k)=c\frac{\lambda^k}{k!}$，$k=1,2,\cdots,\lambda>0$。

解：(1) $1=\sum_{k=1}^{N}f(k)=\frac{c}{N}\cdot N$，　　$\therefore c=1$。

(2) $1=c\sum_{k=1}^{\infty}\frac{\lambda^k}{k!}=c(e^\lambda-1)$，　　$\therefore c=(e^\lambda-1)^{-1}$。

4．证明函数 $f(x)=\frac{1}{2}e^{-|x|}(-\infty<x<\infty)$ 是一个密度函数。

证：$f(x)\geqslant 0$，且

$$\int_{-\infty}^{\infty}f(x)dx=\int_{-\infty}^{\infty}\frac{1}{2}e^{-|x|}dx=\int_0^{\infty}e^{-x}dx=-e^{-x}\Big|_0^{\infty}$$
$$=1,$$

$\therefore$ $f(x)$是一个密度函数。

5．若 ξ 的分布函数为$N(10,4)$，求 ξ 落在下列范围的概率：(1) $(6,9)$；(2) $(7,12)$；(3) $(13,15)$。

解：(1) $P\{6<\xi<9\}=$

$=P\{\frac{1}{2}(6-10)<\frac{1}{2}(\xi-10)<\frac{1}{2}(9-10)\}$

$=P\{-2<\frac{1}{2}(\xi-10)<-\frac{1}{2}\}=$

$=\Phi(-\frac{1}{2})-\Phi(-2)=0.285788$。

(2) $P\{7<\xi<12\}=P\{\frac{1}{2}(7-10)<\frac{1}{2}(\xi-10)<\frac{1}{2}(12-10)\}$

$=P\{-1\frac{1}{2}<\frac{1}{2}(\xi-10)<1\}=\Phi(1)-\Phi(-1\frac{1}{2})$

$=0.774538.$

(3) $P\{13<\xi<15\}=$

$=P\{\frac{1}{2}(13-10)<\frac{1}{2}(\xi-10)<\frac{1}{2}(15-10)\}$

$=P\{1\frac{1}{2}<\frac{1}{2}(\xi-10)<2\frac{1}{2}\}=\Phi(2\frac{1}{2})-\Phi(1\frac{1}{2})$

$=0.060597.$

6. 若 ξ 的分布函数为 $N(60,9)$，求分点 x_1,x_2,x_3,x_4，使 ξ 落在 $(-\infty,x_1),(x_1,x_2),(x_2,x_3),(x_3,x_4),(x_4,\infty)$ 中的概率之比为7:24:38:24:7。

解： $7+24+38+24+7=100$，$P\{\xi<x_4\}=(100-7)/100=0.93$，$P\{\xi<x_3\}=(7+24+38)/100=0.69$。查表得 $\Phi(1.5)\approx0.93$，$\Phi(0.5)\approx0.69$。由题设得

$\Phi(x)=P\{\frac{1}{3}(\xi-60)<\frac{1}{3}(y-60)=x\}=P\{\xi<y\}$

令 $x=\frac{1}{3}(y-60)=1.5$，解得 $y=64.5$，即 $x_4=64.5$。由对称性得 $x_1=60-(64.5-60)=55.5$。再令 $\frac{1}{3}(y-60)=0.5$，解得 $y=61.5$，即 $x_3=61.5$。由对称性得 $x_2=60-(61.5-60)=58.5$。

7. 若 ξ 的分布函数为 $N(5,4)$，求 a 使：

(1) $P\{\xi<a\}=0.90$；(2) $P\{|\xi-5|>a\}=0.01$。

解： (1) $\Phi(1.3)=0.90$，而 $P\{\xi<a\}=P\{\frac{1}{2}(\xi-5)<\frac{1}{2}(a-5)\}=\Phi(\frac{1}{2}(a-5))$，令 $\frac{1}{2}(a-5)=1.3$ 解得 $a=7.6$。

(2) 由 $P\{|\xi-5|>a\}=0.01$ 得 $P\{\xi-5>a\}=0.005$，从而 $P\{\frac{1}{2}(\xi-5)\leqslant\frac{1}{2}a\}=0.995$，而 $\Phi(2.6)=0.995$ 所以

$\frac{1}{2}a=2.6$，$a=5.2$。

8. 设 $F(x)=P\{\xi\leqslant x\}$，试证 $F(x)$ 具有下列性质：

(1) 非降；(2) 右连续；(3) $F(-\infty)=0$，$F(+\infty)=1$。

证：(1) 设 $x_2>x_1$，$F(x_2)-F(x_1)=P\{x_1<\xi\leqslant x_2\}\geqslant 0$，所以 $F(x_2)\geqslant F(x_1)$，$F(x)$ 非降。

(2) 设 $x<\cdots<x_n<x_{n-1}<\cdots<x_1<x_0$，$x_i\downarrow x$。由概率的可加性得

$$P\{\bigcup_{i=0}^{\infty}(x_{i+1}<\xi\leqslant x_i)\}=P\{x<\xi\leqslant x_0\},$$

$$\sum_{i=0}^{\infty}[F(x_i)-F(x_{i+1})]=F(x_0)-F(x).$$

由此得　$F(x_0)-F(x)=\lim\limits_{n\to\infty}[F(x_0)-F(x_n)]$，

$\therefore$　$F(x)=\lim\limits_{n\to\infty}F(x_n)=F(x+0)$，$F(x)$ 右连续。

(3)　$1=P\{-\infty<\xi<\infty\}=\sum\limits_{n=-\infty}^{\infty}P\{n<\xi\leqslant n+1\}$

$$=\sum_{n=-\infty}^{\infty}[F(n+1)-F(n)]=\lim_{n\to\infty}F(n)-\lim_{m\to-\infty}F(m).$$

由单调性得 $\lim\limits_{x\to-\infty}F(x)$ 与 $\lim\limits_{x\to\infty}F(x)$ 均存在且有穷，由 $0\leqslant F(x)\leqslant 1$ 及上式得 $F(-\infty)=0$，$F(\infty)=1$。

9. 试证：若 $P\{\xi\leqslant x_2\}\geqslant 1-\beta$，$P\{\xi\geqslant x_1\}\geqslant 1-\alpha$，则 $P\{x_1\leqslant\xi\leqslant x_2\}\geqslant 1-(\alpha+\beta)$。

证：
$$\begin{aligned}P\{x_1\leqslant\xi\leqslant x_2\}&=P\{\xi\leqslant x_2\}-P\{\xi<x_1\}\\&=P\{\xi\leqslant x_2\}-(1-P\{\xi\geqslant x_1\})\\&=P\{\xi\leqslant x_2\}+P\{\xi\geqslant x_1\}-1\\&\geqslant(1-\beta)+(1-\alpha)-1=1-(\alpha+\beta).\end{aligned}$$

$\therefore$ 不等式成立。

10. 设随机变量 ξ 取值于 $[0,1]$，若 $P\{x\leqslant\xi<y\}$ 只与长度 $y-x$ 有关（对一切 $0\leqslant x\leqslant y\leqslant 1$），试证 ξ 服从 $[0,1]$ 均匀分布。

证法一：定义

$$F(x)=\begin{cases}0, & x\in(-\infty,0]\\ P\{0\leqslant\xi<x\}, & x\in(0,1]\\ 1, & x\in(1,\infty)\end{cases}$$

则$F(x)$是ξ的分布函数。由题设得，对任意$2x\in[0,1]$有，$P\{0\leqslant\xi<x\}=P\{x\leqslant\xi<2x\}$，即有$P\{0\leqslant\xi<2x\}=2P\{0\leqslant\xi<x\}$。由此得$F(2x)=2F(x)$。逐一类推可得，若$nx\in[0,1]$，则$F(nx)=nF(x)$，或者$\frac{1}{n}F(x)=F\left(\frac{x}{n}\right)$。从而对有理数$\frac{m}{n}$，若$\frac{m}{n}x$与$x$都属于$[0,1]$，则有$F(\frac{m}{n}x)=\frac{m}{n}F(x)$。再由$F(x)$的左连续性可得，对任意无理数$\alpha$，若$\alpha x$与$x$都属于$[0,1]$，则$F(\alpha x)=\alpha F(x)$。

因为区间$[0,1)$与$[0,1]$的长度相等，由题设得

$F(1)=P\{0\leqslant\xi<1\}=P\{0\leqslant\xi\leqslant1\}=1$。

由此及上段证明得，对任意$x\in[0,1]$有$F(x)=xF(1)=x$，即$F(x)$为

$$F(x)=\begin{cases}0, & x\leqslant0\\ x, & 0<x\leqslant1\\ 1, & x>1\end{cases}$$

$\therefore$ ξ服从$[0,1]$上均匀分布。

证法二：如同证法一中定义ξ的分布函数$F(x)$。由$F(x)$单调知它对$[0,1]$上的L-测度几乎处处可微。设$x_1,x_2\in(0,1)$，当$x_i+\Delta x\in[0,1]$（$i=1,2$）时，由题设得

$$\begin{aligned}F(x_1+\Delta x)-F(x_1)&=P\{x_1\leqslant\xi<x_1+\Delta x\}\\&=P\{x_2\leqslant\xi<x_2+\Delta x\}=F(x_2+\Delta x)-F(x_2)\end{aligned}$$

等式两端都除以Δx，再令$\Delta x\to0$可得，由$F'(x_1)$存在可推得$F'(x_2)$也存在，而且$F'(x_2)=F'(x_1)$。从而对任意$x\in(0,1)$有$F'(x)\equiv c$。当$x\overline{\in}[0,1]$时显然有$F'(x)=0$。一

点的长度为0，由题设得$P\{\xi=0\}=P\{\xi=1\}=0$。由上所述可知ξ是连续型随机变量，$F'(x)$是其密度函数，从而定出$c=1$。至此得证ξ服从$[0,1]$均匀分布。

*11. 若存在Θ上的实值函数$Q(\theta)$及$D(\theta)$以及$T(x)$及$S(x)$，使

$$f_\theta(x)=\exp\{Q(\theta)T(x)+D(\theta)+S(x)\},$$

则称$\{f_\theta,\ \theta\in\Theta\}$是一个**单参数的指数族**。证明 (1) 正态分布$N(m_0,\sigma^2)$，已知m_0，关于参数σ；(2) 正态分布$N(m,\sigma_0^2)$，已知σ_0，关于参数m；(3) 普阿松分布$p(k;\lambda)$关于λ都是一个单参数的指数族。

但$[0,\theta]$上的均匀分布，关于θ不是一个单参数的指数族。

证： (1) $f_\sigma(x)=\dfrac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\dfrac{(x-m_0)^2}{2\sigma^2}\right\}$

$$=\exp\left\{-\frac{1}{2\sigma^2}(x-m_0)^2+\ln\frac{1}{\sqrt{2\pi}\sigma}\right\}$$

$$=\exp\left\{-\frac{(x-m_0)^2}{2\sigma^2}-\ln\sigma-\ln\sqrt{2\pi}\right\}$$

若令$Q(\sigma)=-1/(2\sigma^2)$，$T(x)=(x-m_0)^2$，$D(\sigma)=-\ln\sigma$，$S(x)=-\ln\sqrt{2\pi}$，则有

$$f_\sigma(x)=\exp\{Q(\sigma)T(x)+D(\sigma)+S(x)\}.$$

这就证明了正态分布$N(m_0,\sigma^2)$是单参数$\sigma(\sigma>0)$的指数族。

(2) $f_m(x)=\dfrac{1}{\sqrt{2\pi}\sigma_0}\exp\left\{-\dfrac{(x-m)^2}{2\sigma_0^2}\right\}$

$$=\frac{1}{\sqrt{2\pi}\sigma_0}\exp\left\{-\frac{x^2-2mx+m^2}{2\sigma_0^2}\right\}$$

$$=\exp\left\{\frac{mx}{\sigma_0^2}-\frac{m^2}{2\sigma_0^2}-\frac{x^2}{2\sigma_0^2}+\ln\frac{1}{\sqrt{2\pi}\sigma_0}\right\}$$

若令$Q(m)=m/\sigma_0^2$，$T(x)=x$，$D(m)=-\frac{1}{2}m^2/\sigma_0^2$，

$S(x)=-\frac{x^2}{2\sigma_0^2}+\ln\frac{1}{\sqrt{2\pi}\sigma_0}$，则

$$f_m(x)=\exp\{Q(m)T(x)+D(m)+S(x)\},$$

所以正态分布$N(m,\ \sigma_0^2)$是单参数$m(-\infty<m<\infty)$的指数族。

(3) $p(k;\lambda)=\frac{\lambda^k}{k!}e^{-\lambda}=\exp\{k\ln\lambda-\lambda-\ln k!\}$.

若令$Q(\lambda)=\ln\lambda$，$T(k)=k$，$D(\lambda)=-\lambda$，$S(k)=-\ln k!$，

则 $p(k;\lambda)=\exp\{Q(\lambda)T(k)+D(\lambda)+S(k)\}$，

所以$p(k;\lambda)$是单参数$\lambda(\lambda>0)$的指数族。

(4)关于$[0,\theta]$上的均匀分布，其密度函数为

$$f_\theta(x)=\begin{cases}1/\theta, & 0\leqslant x\leqslant\theta\\ 0, & x>\theta \text{或} x<0\end{cases}$$

$f_\theta(x)$是定义在$-\infty<x<\infty$的函数，由于它是x的分段表示的函数，所以无法写成形式

$$f_\theta(x)=\exp\{Q(\theta)T(x)+D(\theta)+S(x)\},$$

故$f_\theta(x)$关于θ不是一个单参数的指数族。

12.定义二元函数

$$F(x,y)=\begin{cases}1, & x+y>0\\ 0, & x+y\leqslant 0.\end{cases}$$

验证此函数对每个变元非降，左连续，且满足(2.6)及(2.7)，但无法使(2.5)保持非负。

证：分别对固定的x_0和y_0有

$$F(x_0,y)=\begin{cases}1, & y>-x_0\\ 0, & y\leqslant -x_0,\end{cases}\quad F(x,y_0)=\begin{cases}1, & x>-y_0.\\ 0, & x\leqslant -y_0.\end{cases}$$

由上式显然可得$F(x,y)$对每个变元非降，左连续，而且满足

(2.6)及(2.7)，即$F(-\infty,y)=0$，$F(x,-\infty)=0$，$F(+\infty,+\infty)=1$。但有

$$F(1,1)-F(1,0)-F(0,1)+F(0,0)=-1,$$

这说明当取$a_1=a_2=0$，$b_1=b_2=1$时(2.5)式不成立。所以$F(x,y)$不是分布函数。

13. 试证$f(x,y)=ke^{-(ax^2+2bxy+cy^2)}$为密度函数的充要条件为$a>0, c>0, b^2-ac<0, k=\sqrt{ac-b^2}/\pi$。

证： 必要性。

$$\iint f(x,y)dxdy=$$

$$=\iint ke^{-a\left(x+\frac{b}{a}y\right)^2}\cdot e^{-\frac{ac-b^2}{a}y}\,dxdy.$$

令$u=x+\frac{b}{a}y$，$v=y$，得$y=v$，$x=u-\frac{b}{a}v$，$J=1$。设

$$\iint f(x,y)dxdy=\int_{-\infty}^{\infty}ke^{-au^2}du\int_{-\infty}^{\infty}e^{-\frac{ac-b^2}{a}v^2}dv$$

$$=1.$$

要积分收敛，必须$a>0$，$(ac-b^2)/a>0$，由此得应有$ac-b^2>0$以及$c>0$。利用$\int_{-\infty}^{\infty}e^{-u^2}du=\sqrt{\pi}$可得

$$\int_{-\infty}^{\infty}ke^{-au^2}du\int_{-\infty}^{\infty}e^{-\frac{ac-b^2}{a}v^2}dv=$$

$$=k\cdot\frac{1}{\sqrt{a}}\sqrt{\pi}\cdot\frac{\sqrt{a}}{\sqrt{ac-b^2}}\sqrt{\pi}=1.$$

$\therefore \quad k=\sqrt{ac-b^2}/\pi$

从而题中所列条件全部满足。

以上诸步可逆推，充分性显然。

14．若$f_1(x)$，$f_2(y)$为分布密度，求为使$f(x,y)=f_1(x)f_2(y)+h(x,y)$成为密度函数，$h(x,y)$必须而且只需满足什么条件。

解：设$f(x,y)=f_1(x)f_2(y)+h(x,y)$是密度函数，则由$f(x,y)\geqslant 0$得$h(x,y)\geqslant -f_1(x)f_2(y)$。又

$$1=\iint f(x,y)\,dxdy=\int f_1(x)dx\int f_2(y)dy+$$
$$+\iint h(x,y)dxdy=1+\iint h(x,y)dxdy,$$

所以应有$\iint h(x,y)dxdy=0$。

反之，若$h(x,y)\geqslant -f_1(x)f_2(y)$，$h(x,y)$可积且$\iint h(x,y)dxdy=0$，显然有$f(x,y)\geqslant 0$且$\iint f(x,y)dxdy=1$，即$f(x,y)$是密度函数。

所以为使$f(x,y)$是密度函数，$h(x,y)$必须而且只需满足$h(x,y)\geqslant -f_1(x)f_2(y)$且$\iint h(x,y)dxdy=0$。

15．若(ξ,η)的密度函数为

$$f(x,y)=\begin{cases}Ae^{-(2x+y)}, & x>0,\ y>0\\ 0, & \text{其它}\end{cases}$$

试求：(1)常数A；(2)$P\{\xi<2,\eta<1\}$；(3)ξ的边际分布；(4)$P\{\xi+\eta<2\}$；(5)$f(x|y)$；(6)$P\{\xi<2|\eta<1\}$。

解：(1)
$$1=\int_0^{\infty}Ae^{-2x}dx\int_0^{\infty}e^{-y}dy$$
$$=A\left(-\tfrac{1}{2}e^{-2x}\right)\Big|_0^{\infty}\cdot\left(-e^{-y}\Big|_0^{\infty}\right)=A/2,\ A=2$$

(2) $P\{\xi<2,\eta<1\}=\int_0^2 2e^{-2x}dx\int_0^1 e^{-y}dy$

$$=\left(-e^{-2x}\Big|_0^2\right)\left(-e^{-y}\Big|_0^1\right)=(1-e^{-4})(1-e^{-1}).$$

(3) ξ的边际分布，当$x\leqslant 0$时$f_\xi(x)=0$；当$x>0$时有

$$f_\xi(x)=\int_0^{\infty}2e^{-2x}e^{-y}dy=2e^{-2x}.$$

(4) $P\{\xi+\eta<2\}=\int_0^2 2e^{-2x}dx\int_0^{2-x}e^{-y}dy$

$$=\int_0^2 2e^{-2x}(1-e^{-(2-x)})dx=\int_0^2(2e^{-2x}-2e^{-(2+x)})dx$$

$$=(1-e^{-4})+(2e^{-4}-2e^{-2})=1+e^{-4}-2e^{-2}=(1-e^{-2})^2.$$

(5) 当$x<0$，$y>0$时$f(x|y)=0$；当$x>0$，$y>0$时有

$$f(x|y)=\frac{f(x,y)}{f_\eta(y)}=\frac{2e^{-(2x+y)}}{e^{-y}}=2e^{-2x}.$$

(6) $P\{\eta<1\}=\int_0^1 dy\int_0^{\infty}2e^{-(2x+y)}dx$

$$=\int_0^1 e^{-y}dy\int_0^{\infty}2e^{-2x}dx$$

$$=-e^{-y}\Big|_0^1=1-e^{-1},$$

利用(2)的结果可得

$$P\{\xi<2|\eta<1\}=\frac{P\{\xi<2,\eta<1\}}{P\{\eta<1\}}=\frac{(1-e^{-4})(1-e^{-1})}{1-e^{-1}}$$

$$=1-e^{-4}.$$

16. 若(ξ,η)服从二元正态分布，参数$a,b,\sigma_1^2,\sigma_2^2,r$，以$D(\lambda)$记下面椭圆的内部

$$\frac{(x-a)^2}{\sigma_1^2}-\frac{2r(x-a)(y-b)}{\sigma_1\sigma_2}+\frac{(y-b)^2}{\sigma_2^2}=\lambda^2$$

试求$P\{(\xi,\eta)\in D(\lambda)\}$。

解： 作变换，令$x-a=\rho\cos\theta$，$y-b=\rho\sin\theta$，则$|J|=\rho$，

椭园区域为

$$\rho^2\left\{\frac{\cos^2\theta}{\sigma_1^2}-\frac{2r\sin\theta\cos\theta}{\sigma_1\sigma_2}+\frac{\sin^2\theta}{\sigma_2^2}\right\}=\lambda^2$$

记 $$\frac{\cos^2\theta}{\sigma_1^2}-\frac{2r\sin\theta\cos\theta}{\sigma_1\sigma_2}+\frac{\sin^2\theta}{\sigma_2^2}=s^2$$

则$\rho=\lambda/s$，且

$$P\{(\xi,\eta)\in D(\lambda)\}=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}}\times$$

$$\times\int_0^{2\pi}d\theta\int_0^{\frac{\lambda}{S}}e^{-\frac{1}{2(1-r^2)}\rho^2S^2}\rho d\rho$$

$$=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}}\times$$

$$\times\int_0^{2\pi}-\frac{(1-r^2)}{S^2}e^{-\frac{S^2}{2(1-r^2)}\rho^2}\Big|_0^{\frac{\lambda}{S}}d\theta$$

$$=\frac{\sqrt{1-r^2}}{2\pi\sigma_1\sigma_2}\left(1-e^{-\frac{\lambda^2}{2(1-r^2)}}\right)\int_0^{2\pi}\frac{1}{S^2}d\theta$$

当$\lambda\to\infty$时，$P\{(\xi,\eta)\in D(\lambda)\}\to 1$，由此得

$$\int_0^{2\pi}\frac{1}{S^2}d\theta=\frac{2\pi\sigma_1\sigma_2}{\sqrt{1-r^2}}$$

$$\therefore\quad P\{(\xi,\eta)\in D(\lambda)\}=1-\exp\left\{-\frac{\lambda^2}{2(1-r^2)}\right\}$$

17.证明多项分布的边际分布仍是多项分布。

证： 设多项分布为

$$P\{\xi_1=k_1,\cdots,\xi_r=k_r\}=\frac{n!}{k_1!\cdots k_r!}p_1^{k_1}\cdots p_r^{k_r},\quad(1)$$

$$k_i\geqslant 0,\ \sum_{i=1}^{r}k_i=n,\ \sum_{i=1}^{r}p_i=1.\qquad(2)$$

利用(2)可以把(1)改写成

$$P\{\xi_1=k_1\cdots,\xi_{r-1}=k_{r-1}\}=$$

$$=\frac{n!}{k_1!\cdots k_{r-1}!(n-k_1-\cdots-k_{r-1})!}p_1^{k_1}\cdots p_{r-1}^{k_{r-1}}\times$$

$$\times(1-p_1-\cdots-p_{r-1})^{n-k_1-\cdots-k_{r-1}} \qquad (3)$$

由边际分布的定义并把(3)代入得

$$P\{\xi_1=k_1,\cdots,\xi_{r-2}=k_{r-2}\}$$

$$=\sum_{\substack{k_{r-1}\\ k_1+\cdots+k_{r-1}\leqslant n,k_{r-1}\geqslant 0}} P\{\xi_1=k_1,\cdots,\xi_{r-1}=k_{r-1}\}$$

$$=\frac{n!p_1^{k_1}\cdots p_{r-2}^{k_{r-2}}}{k_1!\cdots k_{r-2}!(n-k_1-\cdots-k_{r-2})!}\times$$

$$\times\sum_{k_{r-1}=0}^{n-k_1-\cdots-k_{r-2}}\frac{(n-k_1-\cdots-k_{r-2})!}{k_{r-1}!(n-k_1-\cdots-k_{r-1})!}p_{r-1}^{k_{r-1}}\times$$

$$\times(1-p_1-\cdots-p_{r-2}-p_{r-1})^{n-k_1-\cdots-k_{r-1}}$$

由二项式定理得

$$P\{\xi_1=k_1,\cdots,\xi_{r-2}=k_{r-2}\}=$$

$$=\frac{n!}{k_1!\cdots k_{r-2}!(n-k_1-\cdots-k_{r-2})!}p_1^{k_1}\cdots p_{r-2}^{k_{r-2}}\times$$

$$\times(1-p_1-\cdots-p_{r-2})^{n-k_1-\cdots-k_{r-2}} \qquad (4)$$

把(4)与(3)比较知，边际分布仍服从多项分布。多次类推可得

$$P\{\xi_1=k_1\}=\frac{n!}{k_1!(n-k_1)!}p_1^{k_1}(1-p_1)^{n-k_1}$$

从而知任意边际分布均服从多项分布(包括二项分布)

18.设二维随机变量(ξ,η)的联合密度为

$$p(x,y)=\frac{1}{\Gamma(k_1)\Gamma(k_2)}x^{k_1-1}(y-x)^{k_2-1}e^{-y}$$

$k_1>0, k_2>0, 0<x\leqslant y<\infty$，试求 ξ 与 η 的边际分布。

解： (1) ξ 的密度函数为，当 $x\leqslant 0$ 时 $p_\xi(x)=0$；当 $x>0$ 时，注意积分限的选取，得

$$p_\xi(x)=\int_{-\infty}^{\infty}p(x,y)dy=\int_x^{\infty}\frac{1}{\Gamma(k_1)\Gamma(k_2)}\times$$

$$\times x^{k_1-1}(y-x)^{k_2-1}e^{-y}dy(\text{令}y-x=t)$$

$$=\frac{x^{k_1-1}}{\Gamma(k_1)\Gamma(k_2)}\int_0^{\infty}t^{k_2-1}e^{-x}e^{-t}dt$$

$$=\frac{x^{k_1-1}}{\Gamma(k_1)}e^{-x}$$

(2) η 的密度函数为，当 $y\leqslant 0$ 时 $p_\eta(y)=0$；当 $y>0$ 时，

$$p_\eta(y)=\int_{-\infty}^{\infty}p(x,y)dx=$$

$$=\int_0^y\frac{1}{\Gamma(k_1)\Gamma(k_2)}x^{k_1-1}(y-x)^{k_2-1}e^{-y}dx$$

令 $x=yt$，当 $x=0$ 时 $t=0$，当 $x=y$ 时 $t=1$，所以

$$p_\eta(y)=\frac{e^{-y}}{\Gamma(k_1)\Gamma(k_2)}y^{k_1-1}y^{k_2-1}\times$$

$$\times\int_0^1 t^{k_1-1}(1-t)^{k_2-1}ydt$$

$$=\frac{y^{k_1+k_2-1}e^{-y}}{\Gamma(k_1)\Gamma(k_2)}\cdot B(k_1,k_2)$$

$$=\frac{y^{k_1+k_2-1}e^{-y}}{\Gamma(k_1)\Gamma(k_2)}\cdot\frac{\Gamma(k_1)\Gamma(k_2)}{\Gamma(k_1+k_2)}$$

$$=\frac{1}{\Gamma(k_1+k_2)}y^{k_1+k_2-1}e^{-y}$$

其中用到β-函数与Γ-函数的关系式。

19．若$f_1(x),f_2(x),f_3(x)$是对应于分布函数$F_1(x)$，$F_2(x),F_3(x)$的密度函数，证明对于一切$\alpha(-1<\alpha<1)$，下列函数是密度函数，且具有相同的边际密度函数$f_1(x),f_2(x)$，$f_3(x)$：

$$f_\alpha(x_1,x_2,x_3)=f_1(x_1)f_2(x_2)f_3(x_3)\{1+\alpha[2F_1(x_1)-1]\times$$
$$\times[2F_2(x_2)-1]\cdot[2F_3(x_3)-1]\}$$

证：我们有

$$0\leqslant F_i(x_i)\leqslant 1,\quad -1\leqslant 2F_i(x_i)-1\leqslant 2-1=1,$$
$$-1\leqslant[2F_1(x_1)-1][2F_2(x_2)-1][2F_3(x_3)-1]\leqslant 1,$$

代入$f_\alpha(x_1,x_2,x_3)$的表达式得

$$f_\alpha(x_1,x_2,x_3)\geqslant 0 \tag{1}$$

又有

$$\int_{-\infty}^{\infty}[2F_i(x_i)-1]f_i(x_i)dx_i=$$
$$=\int_{-\infty}^{\infty}[2F_i(x_i)-1]\ dF_i(x_i)$$
$$=[F_i^2(x_i)-F_i(x_i)]\Big|_{-\infty}^{\infty}=0$$

$$\therefore\iiint f_\alpha(x_1,x_2,x_3)dx_1dx_2dx_3=$$
$$=\int_{-\infty}^{\infty}f_1(x_1)dx_1\int_{-\infty}^{\infty}f_2(x_2)dx_2\int_{-\infty}^{\infty}f_3(x_3)dx_3$$
$$=1 \tag{2}$$

由（1），（2）知$f_\alpha(x_1,x_2,x_3)$是密度函数。用与上面类似的方法计算可得边际密度函数为

$$\iint f_\alpha(x_1,x_2,x_3)dx_2dx_3=f_1(x_1),$$

$$\iint f_\alpha(x_1,x_2,x_3)dx_1dx_2 = f_3(x_3),$$

$$\iint f_\alpha(x_1,x_2,x_3)dx_1dx_3 = f_2(x_2)$$

20. 设ξ与η是相互独立的随机变量，均服从几何分布

$$g(k,p) = q^{k-1}p,\quad k = 1,2,\cdots.$$

令$\zeta = \max(\xi,\eta)$，试求(1)(ζ,ξ)的联合分布；(2)ζ的分布；(3)ξ关于ζ的条件分布

解： (1)为求(ζ,ξ)的联合概率分布，分别考虑下列三种情况：$(i,k\geqslant 1)$其中利用到独立性，

(a)$i = k$，

$$P\{\zeta = k,\xi = k\} = P\{\bigcup_{j=1}^{k}(\xi = k,\eta = j\} = \sum_{j=1}^{k}P\{\xi = k,\eta = j\}$$

$$= \sum_{j=1}^{k} p^2q^{k+j-2} = p^2q^{k-1}\cdot\frac{1-q^k}{1-q} = pq^{k-1}(1-q^k)$$

(b)$i<k$，

$P\{\zeta = k,\xi = i\} = P\{\xi = i,\eta = k\} = p^2q^{i+k-2}$；

(c)$i>k$，

$$\{\zeta = k,\xi = i\} = \phi, P\{\zeta = k,\xi = i\} = 0$$

(2)因为$\zeta = \max(\xi,\eta)$，所以

$$\{\zeta = k\} = \bigcup_{i=1}^{k-1}\{\xi = i,\eta = k\}\cup\bigcup_{j=1}^{k}\{\xi = k,\eta = j\}$$

$$P\{\zeta = k\} = \sum_{i=1}^{k-1}P\{\xi = i,\eta = k\} + \sum_{j=1}^{k}P\{\xi = k,\eta = j\}$$

$$= \sum_{i=1}^{k-1}p^2q^{i+k-2} + \sum_{j=1}^{k}p^2q^{k+j-2} =$$

$$=p^2q^{k-1}\left(\frac{1-q^{k-1}}{1-q}+\frac{1-q^k}{1-q}\right)=(2-q^{k-1}-q^k)pq^{k-1}$$

$$(k=1,2,\cdots)$$

(3) $P\{\xi=i|\zeta=k\}=\dfrac{P\{\xi=i,\zeta=k\}}{P\{\zeta=k\}}$

$$=\begin{cases}\dfrac{pq^{k-1}(1-q^k)}{pq^{k-1}(2-q^{k-1}-q^k)}=\dfrac{1-q^k}{2-q^{k-1}-q^k}, & i=k\\ \dfrac{p^2q^{i+k-2}}{pq^{k-1}(2-q^{k-1}-q^k)}=\dfrac{pq^{i-1}}{2-q^{k-1}-q^k}, & i<k\\ 0\quad, & i>k,(i,k\geqslant1)\end{cases}$$

21.(1)若(ξ,η)的联合密度函数为

$$f(x,y)=\begin{cases}4xy, & 0\leqslant x\leqslant1,\ 0\leqslant y\leqslant1\\ 0, & 其它,\end{cases}$$

问ξ与η是否相互独立

(2)若(ξ,η)的联合密度函数为

$$g(x,y)=\begin{cases}8xy, & 0\leqslant x\leqslant y,\ 0\leqslant y\leqslant1\\ 0, & 其它\end{cases}$$

问ξ与η是否相互独立?

解:(1)边际分布的密度函数为,当$x\overline{\in}[0,1]$时$f_\xi(x)=0$,当$0\leqslant x\leqslant1$时,

$$f_\xi(x)=\int_{-\infty}^{\infty}f(x,y)dy=\int_0^1 4xydy=2x$$

同理,当$y\overline{\in}[0,1]$时$f_\eta(y)=0$,当$0\leqslant y\leqslant1$时$f_\eta(y)=2y$。$f(x,y)=f_\xi(x)f_\eta(y)$,所以ξ与η独立。

(2)边际密度函数为,当$x\overline{\in}[0,1]$时$f_\xi(x)=0$;当$0<x<1$时

$$f_\xi(x)=\int_{-\infty}^{\infty}g(x,y)\,dy=\int_x^1 8xydy=4x(1-x^2)$$

当$y\overline{\in}[0,1]$时$f_\eta(y)=0$,当$0<y<1$时

$$f_\eta(y)=\int_{-\infty}^{\infty} g(x,y)dx=\int_0^y 8xydx=4y^3$$

在区域$0<x<1, 0<y<1$中均有$g(x,y)\neq f_\xi(x)f_\eta(y)$，所以$\xi$与$\eta$不独立。

22.设(ξ,η,ζ)的联合密度函数为

$$p(x,y,z)=\begin{cases}\dfrac{1}{8\pi^3}(1-\sin x\sin y\sin z), & \text{当}\begin{matrix}0\leqslant x\leqslant 2\pi\\0\leqslant y\leqslant 2\pi\\0\leqslant z\leqslant 2\pi\end{matrix}\text{时}\\ 0, & \text{其它}\end{cases}$$

试证ξ,η,ζ两两独立，但不互相独立。

证： 当$0\leqslant x\leqslant 2\pi$，$0\leqslant y\leqslant 2\pi$时，$\xi$与$\eta$的联合分布密度为

$$p_{\xi\eta}(x,y)=\int_0^{2\pi}\frac{1}{8\pi^3}(1-\sin x\sin y\sin z)dz$$

$$=\left[\frac{z}{8\pi^3}-\sin x\sin y(-\cos z)\right]\Big|_0^{2\pi}=\frac{1}{4\pi^2};$$

其余$p_{\xi\eta}(x,y)=0$。当$0\leqslant x\leqslant 2\pi$时，

$$p_\xi(x)=\int_0^{2\pi}dy\int_0^{2\pi}\frac{1}{8\pi^3}(1-\sin x\sin y\sin z)dz=\frac{1}{2\pi};$$

其余$p_\xi(x)=0$。由于ξ,η,ζ三者在密度函数的表达式中所处地位相同，故得当$0\leqslant x\leqslant 2\pi, 0\leqslant z\leqslant 2\pi$时，$p_{\xi\zeta}(x,z)=1/4\pi^2$；当$0\leqslant y\leqslant 2\pi, 0\leqslant z\leqslant 2\pi$时，$p_{\eta\zeta}(y,z)=1/4\pi^2$；当$0\leqslant y\leqslant 2\pi$时，$p_\eta(y)=1/2\pi$；当$0\leqslant z\leqslant 2\pi$时，$p_\zeta(z)=1/2\pi$；在其余区域内，诸边际密度函数均取0值。由于$p_{\xi\eta}(x,y)=p_\xi(x)p_\eta(y)$，$p_{\xi\zeta}(x,z)=p_\xi(x)p_\zeta(z)$，$p_{\eta\zeta}(y,z)=p_\eta(y)p_\zeta(z)$，故$\xi,\eta,\zeta$两两独立；但当$0<x<2\pi, 0<y<2\pi, 0<z<2\pi$时有$p(x,y,z)\neq p_\xi(x)p_\eta(y)p_\zeta(z)$，故$\xi,\eta,\zeta$不相互独立。

23.设(ξ,η)具有联合密度函数

$$p(x,y)=\begin{cases}\dfrac{1+xy}{4}, & |x|<1,\quad |y|<1\\ 0, & \text{其它。}\end{cases}$$

试证ξ与η不独立，但ξ^2与η^2是相互独立的

证： 当$|x|<1$时，

$$p_\xi(x)=\int_{-\infty}^{\infty}p(x,y)dy=\int_{-1}^{1}\frac{1+xy}{4}dy=\frac{1}{2},$$

其余$p_\xi(x)=0$。同理当$|y|<1$时，$p_\eta(y)=\frac{1}{2}$，其余$p_\eta(y)=0$。当$0<|x|<1, 0<|y|<1$时有$p(x,y)\neq p_\xi(x)p_\eta(y)$，所以$\xi$与$\eta$不独立。

现试通过分布函数来证ξ^2与η^2独立。ξ^2的分布函数记为$F_1(x)$，则当$0<x\leqslant 1$时，

$$F_1(x)=P\{\xi^2<x\}=P\{-\sqrt{x}<\xi<\sqrt{x}\}$$

$$=\int_{-\sqrt{x}}^{\sqrt{x}}\frac{1}{2}\,dx=\sqrt{x};$$

同理可求得η^2的分布函数$F_2(y)$，得

$$F_1(x)=\begin{cases}0, & x\leqslant 0\\ \sqrt{x}, & 0<x\leqslant 1\\ 1, & x>1,\end{cases}\qquad F_2(y)=\begin{cases}0, & y\leqslant 0\\ \sqrt{y}, & 0<y\leqslant 1\\ 1, & y>1\end{cases}$$

(ξ^2,η^2)联合分布函数记为$F_3(x,y)$，则当$0\leqslant x\leqslant 1, y\geqslant 1$时

$$F_3(x,y)=P\{\xi^2<x,\eta^2<y\}=P\{\xi^2<x\}=\sqrt{x}$$

同理得当$0\leqslant y\leqslant 1, x\geqslant 1$时$F_3(x,y)=\sqrt{y}$；当$0\leqslant x\leqslant 1$，$0\leqslant y\leqslant 1$时

$$F_3(x,y)=P\{\xi^2<x,\eta^2<y\}$$

$$=P\{-\sqrt{x}<\xi<\sqrt{x},\ -\sqrt{y}<\eta<\sqrt{y}\}$$

$$=\int_{-\sqrt{x}}^{\sqrt{x}}ds\int_{-\sqrt{y}}^{\sqrt{y}}\frac{1+st}{4}\,dt=\sqrt{xy}$$

合起来写得

$$F_3(x,y)=\begin{cases}0, & x\leqslant 0\text{或}y\leqslant 0\\ \sqrt{x}, & 0\leqslant x\leqslant 1,\ y\geqslant 1\\ \sqrt{y}, & 0\leqslant y\leqslant 1,\ x\geqslant 1\\ \sqrt{xy}, & 0\leqslant x\leqslant 1,\ 0\leqslant y\leqslant 1\\ 1, & x\geqslant 1,\ y\geqslant 1.\end{cases}$$

不难验证$F_3(x,y)=F_1(x)F_2(y)$对所有x,y都成立，所以ξ^2与η^2独立。

24.若ξ_1与ξ_2是独立随变量，均服从普阿松分布，参数为λ_1及λ_2，试直接证明

(1)$\xi_1+\xi_2$具有普阿松分布，参数为$\lambda_1+\lambda_2$；

(2)$P\{\xi_1=k|\xi_1+\xi_2=n\}=$

$$=\binom{n}{k}\left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^k\left(\frac{\lambda_2}{\lambda_1+\lambda_2}\right)^{n-k}$$

证：(1)由褶积公式及独立性得

$$P\{\xi_1+\xi_2=k\}=\sum_{i=0}^{k}P\{\xi_1=i,\ \xi_2=k-i\}$$

$$=\sum_{i=0}^{k}P\{\xi_1=i\}P\{\xi_2=k-i\}$$

$$=\sum_{i=0}^{k}\frac{\lambda_1^i}{i!}e^{-\lambda_1}\cdot\frac{\lambda_2^{k-i}}{(k-i)!}e^{-\lambda_2}$$

$$=\frac{1}{k!}e^{-(\lambda_1+\lambda_2)}\sum_{i=0}^{k}\frac{k!}{i!(k-i)!}\lambda_1^i\lambda_2^{k-i}$$

$$=\frac{(\lambda_1+\lambda_2)^k}{k!}e^{-(\lambda_1+\lambda_2)}.\quad k=0,1,2,\cdots$$

这就证明了$\xi_1+\xi_2$具有普阿松分布，且参数为$\lambda_1+\lambda_2$。

(2) $P\{\xi_1=k|\xi_1+\xi_2=n\}=\dfrac{P\{\xi_1=k,\xi_1+\xi_2=n\}}{P\{\xi_1+\xi_2=n\}}$

$$=\frac{P\{\xi_1=k,\xi_2=n-k\}}{P\{\xi_1+\xi_2=n\}}=\frac{P\{\xi_1=k\}P\{\xi_2=n-k\}}{P\{\xi_1+\xi_2=n\}}$$

$$=\frac{\lambda_1^k}{k!}e^{-\lambda_1}\cdot\frac{\lambda_2^{n-k}}{(n-k)!}e^{-\lambda_2}\div$$

$$\div\frac{(\lambda_1+\lambda_2)^n}{n!}e^{-(\lambda_1+\lambda_2)}$$

$$=\binom{n}{k}\left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^k\left(\frac{\lambda_2}{\lambda_1+\lambda_2}\right)^{n-k}$$ 证毕。

25. 若ξ,η相互独立，且皆以概率$\frac{1}{2}$取值$+1$及-1，令$\zeta=\xi\eta$，试证ξ,η,ζ两两独立但不相互独立。

证：由题设得

$P\{\zeta=1\}=P(\{\xi=1,\eta=1\}\cup\{\xi=-1,\eta=-1\})$

$=\frac{1}{2}\cdot\frac{1}{2}+\frac{1}{2}\cdot\frac{1}{2}=\frac{1}{2}$,

$P\{\zeta=-1\}=P(\{\xi=1,\eta=-1\}\cup\{\xi=-1,\eta=1\})$

$=\frac{1}{2}\cdot\frac{1}{2}+\frac{1}{2}\cdot\frac{1}{2}=\frac{1}{2}$.

$P\{\xi=1,\zeta=1\}=$

$=P(\{\xi=1\}\cap[\{\xi=1,\eta=1\}\cup\{\xi=-1,\eta=-1\}])$

$=P\{\xi=1,\eta=1\}=P\{\xi=1\}P\{\eta=1\}$

$=\frac{1}{4}=P\{\xi=1\}P\{\zeta=1\}$,

$P\{\xi=1,\zeta=-1\}=$

$=P(\{\xi=1\}\cap[\{\xi=1,\eta=-1\}\cup\{\xi=-1,\eta=1\}])$

$=P\{\xi=1,\eta=-1\}=P\{\xi=1\}P\{\eta=-1\}$

$=\frac{1}{4}=P\{\xi=1\}P\{\zeta=-1\}$,

同理可证$P\{\xi=-1,\zeta=1\}=P\{\xi=-1\}P\{\zeta=1\}$,

$$P\{\xi=-1,\zeta=-1\}=P\{\xi=-1\}P\{\zeta=-1\}.$$

所以 ξ 与 ζ 相互独立。用同样的方法可证 η 与 ζ 也相互独立。

但

$$P\{\xi=1,\eta=1,\zeta=1\}=P(\{\xi=1,\eta=1\}\cap(\{\xi=1,\eta=1\}\cup\{\xi=-1,\eta=-1\}))=P\{\xi=1,\eta=1\}=\frac{1}{4},$$

$$P\{\xi=1\}P\{\eta=1\}P\{\zeta=1\}=\frac{1}{8},$$

所以 ξ,η,ζ 只两两独立而不相互独立。

26. 若 ξ 服从普阿松分布，参数为 λ，试求(1) $\eta=a\xi+b$；(2) $\eta=\xi^2$ 的分布。

解： $P\{\xi=k\}=\frac{\lambda^k}{k!}e^{-\lambda},\ k=0,1,2,\cdots,$

由此得

$$(1)P\{\eta=ak+b\}=\frac{\lambda^k}{k!}e^{-\lambda},\ k=0,1,2,\cdots,$$

$$(2)P\{\eta=k^2\}=\frac{\lambda^k}{k!}e^{-\lambda},\ k=0,1,2,\cdots.$$

27. 设 ξ 的密度函数为 $p(x)$，求下列随机变量的分布函数：(1) $\eta=\xi^{-1}$，这里 $P\{\xi=0\}=0$；(2) $\eta=\operatorname{tg}\xi$；(3) $\eta=|\xi|$。

解： (1)由 $P\{\xi=0\}=0$ 知，η 以概率 1 取有限值。当 $y>0$ 时，

$$F_\eta(y)=P\{1/\xi<y\}=P\{\xi<0\}+P\{\xi>1/y\}$$

$$=\int_{-\infty}^{0}p(x)\,dx+\int_{1/y}^{\infty}p(x)dx;$$

当 $y<0$ 时，

$$F_\eta(y)=P\{1/\xi<y\}=P\{1/y<\xi<0\}=\int_{1/y}^{0}p(x)dx;$$

当 $y=0$ 时，

$$F_\eta(y)=\int_{-\infty}^{0}p(x)dx.$$

(2) $F_\eta(y)=P\{\operatorname{tg}\xi<y\}$

$$=P\left(\bigcup_{k=-\infty}^{\infty}\{k\pi-\frac{\pi}{2}<\xi<k\pi+\operatorname{arctg}y\}\right)$$

$$=\sum_{k=-\infty}^{\infty}\int_{k\pi-\pi/2}^{k\pi+\operatorname{arctg}y}p(x)dx.$$

(3) 当$y\leqslant0$时，$F_\eta(y)=0$；当$y>0$时，

$$F_\eta(y)=P\{|\xi|<y\}=P\{-y<\xi<y\}=\int_{-y}^{y}p(x)dx.$$

28.对圆的直径作近似度量，设其值均匀 分 布 于(a,b)内，试求圆面积的分布密度。

解：设直径为随机变量d，则

$$p_d(x)=\begin{cases}1/(b-a), & a<x<b\\ 0, & 其它.\end{cases}$$

圆面积$S=\frac{1}{4}\pi d^2$。　　当$\frac{1}{4}\pi a^2<y\leqslant\frac{1}{4}\pi b^2$时，

$$F_s(y)=P\{S<y\}=P\{\frac{1}{4}\pi d^2<y\}=P\{d<\sqrt{4y/\pi}\}$$

$$=\int_a^{\sqrt{4y/\pi}}\frac{1}{b-a}dx;$$

当$y\leqslant\frac{1}{4}\pi a^2$时$F_s(y)=0$；当$y>\frac{1}{4}\pi b^2$时$F_s(y)=1$。由此 对$F_s(y)$求导（利用对参变量积分求导法则）得圆面积的分布 密度为，当$y\leqslant\frac{1}{4}\pi a^2$或$y>\frac{1}{4}\pi b^2$时$p_s(y)=0$；当$\frac{1}{4}\pi a^2<y\leqslant\frac{1}{4}\pi b^2$时

$$p_s(y)=F_s'(y)=\frac{\sqrt{\pi y}}{(b-a)\pi y}$$

29.若ξ,η为相互独立的分别服从$[0,1]$均匀分布的随 机变量，试求$\zeta=\xi+\eta$的分布密度函数。

解：ξ与η的密度函数为

$$p_\xi(x)=p_\eta(x)=\begin{cases}1, & 0\leqslant x\leqslant 1\\ 0, & 其它\end{cases} \qquad (1)$$

由卷积公式及独立性得$\zeta=\xi+\eta$的分布密度函数为

$$p_\zeta(y)=\int_{-\infty}^{\infty}p_\xi(x)p_\eta(y-x)dx \qquad (2)$$

把(2)与(1)比较知，在(2)中应有$0\leqslant x\leqslant 1, 0\leqslant y-x\leqslant 1$，满足此不等式组的解$(x,y)$构成图13中平面区域▱ABCD，当$0\leqslant y\leqslant 1$时$0\leqslant x\leqslant y$，当$1\leqslant y\leqslant 2$时$y-1\leqslant x\leqslant 1$。所以当$0\leqslant y\leqslant 1$时(2)中积分为

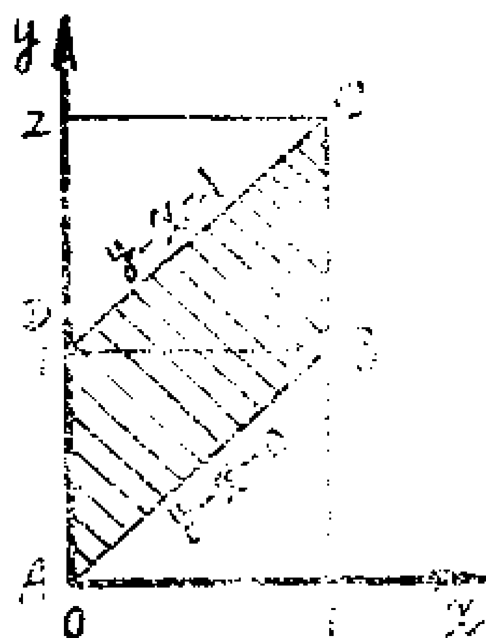


图 13

$$p_\zeta(y)=\int_0^y 1\times 1dx=y$$

当$1\leqslant y\leqslant 2$时，(2)中积分为

$$p_\zeta(y)=\int_{y-1}^{1}1\times 1dx=2-y;$$

对其余的y有$p_\zeta(y)=0$。

30．设ξ,η相互独立，分别服从$N(0,1)$，试求$\zeta=\xi/\eta$的密度函数。

解：$p_\xi(x)=p_\eta(x)=\dfrac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}$

$$p_{\xi\eta}(x,y)=\frac{1}{2\pi}e^{-\frac{1}{2}(x^2+y^2)}$$

由求商的密度函数的公式得

$$p_\zeta(y)=\int_{-\infty}^{\infty}|x|\,p(xy,x)\,dx=$$

$$=\int_{-\infty}^{\infty}|x|\ \frac{1}{2\pi}\ e^{-\frac{1}{2}(x^2y^2+x^2)}\ dx$$

$$=\frac{2}{2\pi}\int_0^{\infty}\ xe^{-\frac{1}{2}x^2(1+y^2)}dx=$$

$$=\frac{1}{\pi}\cdot\frac{1}{1+y^2}\left[-e^{-\frac{1}{2}x^2(1+y^2)}\right]_0^{\infty}$$

$$=\frac{1}{\pi(1+y^2)},\quad -\infty<y<+\infty.$$

$\zeta=\xi/\eta$服从柯西分布。

31.若ξ,η是独立随机变量，均服从$N(0,1)$，试求$U=\xi+\eta$，$V=\xi-\eta$的联合密度函数。

解：作变换，令$s=x+y$，$t=x-y$，得$x=\frac{1}{2}(s+t)$，$y=\frac{1}{2}(s-t)$，$|J|=\frac{1}{2}$。由ξ与η独立知，它们的联合密度应是它们单个密度的乘积，由此得U,V的联合密度函数为

$$p_{UV}(s,t)=\frac{1}{\sqrt{2\pi}}\ e^{-\frac{1}{2}x^2}\cdot\frac{1}{\sqrt{2\pi}}\ e^{-\frac{1}{2}y^2}\cdot|J|=$$

$$=\frac{1}{2\pi}\ e^{-\frac{1}{2}\left[\left(\frac{s+t}{2}\right)^2+\left(\frac{s-t}{2}\right)^2\right]}\cdot\frac{1}{2}$$

$$=\frac{1}{4\pi}\ e^{-\frac{1}{4}(s^2+t^2)}$$

$$=\frac{1}{\sqrt{2\pi\cdot2}}\ e^{-\frac{1}{2}(s/\sqrt{2})^2}\times$$

$$\times\frac{1}{\sqrt{2\pi\cdot2}}\ e^{-\frac{1}{2}(t/\sqrt{2})^2}$$

$$=p_U(s)p_V(t).$$

所以U,V两随机变量也相互独立，且均服从$N(0,2)$。

32.若$\xi_1,\xi_2,\cdots,\xi_n$相互独立，且皆服从指数分布，参

数分别为$\lambda_1,\lambda_2,\cdots,\lambda_n$，试求$\eta=\min(\xi_1,\xi_2,\cdots,\xi_n)$的分布。

解： 当$y>0$时由独立性得（参看教本P151）

$$1-F_\eta(y)=P\{\eta\geqslant y\}=P\{\xi_1\geqslant y,\xi_2\geqslant y,\cdots,\xi_n\geqslant y\}$$

$$=\prod_{i=1}^{n}P\{\xi_i\geqslant y\}=\prod_{i=1}^{n}(1-F_{\xi_i}(y))$$

$$=\prod_{i=1}^{n}(e^{-\lambda_i y})=\exp(-y\sum_{i=1}^{n}\lambda_i)$$

$$\therefore\ F_\eta(y)=1-\exp(-y\sum_{i=1}^{n}\lambda_i)$$

当$y\leqslant 0$时$F_\eta(y)=0$。求导得η的密度函数为，当$y\leqslant 0$时$p_\eta(y)=0$；当$y>0$时

$$p_\eta(y)=F_\eta'(y)=\sum_{j=1}^{n}\lambda_j\exp(-y\sum_{i=1}^{n}\lambda_i)$$

33．在$(0,a)$线段上随机投掷两点，试求两点间距离的分布函数。

解： 设在$(0,a)$内任意投两点ξ_1,ξ_2，其坐标分别为x,y，则ξ_1,ξ_2的联合分布密度为

$$p(x,y)=\begin{cases}0, & (x,y)\overline{\in}(0,a)\times(0,a)\\ \dfrac{1}{a^2}, & (x,y)\in(0,a)\times(0,a).\end{cases}$$

设$\eta=|\xi_1-\xi_2|$，则η的分布函数为，

当$z\leqslant 0$时$F_\eta(z)=0$；当$z>a$时$F_\eta(z)=1$；当$0<z\leqslant a$时，

$$F_\eta(z)=P\{|\xi_1-\xi_2|<z\}$$

$$=\iint\limits_{\substack{-z<x-y<z\\0<x,y<a}}p(x,y)dxdy=\frac{1}{a^2}\iint\limits_{\substack{-z<x-y<z\\0<x,y<a}}dxdy=\frac{1}{a^2}S,$$

积分S为平面区域ABCDEF(图14)的面积，其值为

$a^2-(a-z)^2=2az-z^2$，所以$F_\eta(z)=(2az-z^2)/a^2$

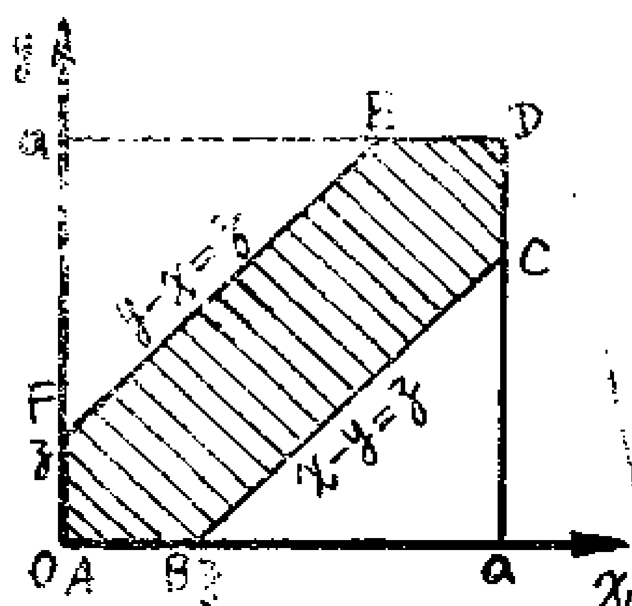


图 14

34. 若气体分子的速度是随机向量 $V=(x,y,z)$，各分量相互独立，且均服从$N=(0,\sigma^2)$，试证 $S=\sqrt{x^2+y^2+z^2}$ 服从马克斯威尔分布

$$f(s)=\sqrt{\frac{2}{\pi}}\frac{s^2}{\sigma^3}\exp\left(-\frac{s^2}{2\sigma^2}\right),\ s>0.$$

证：由独立性得，$V=(x,y,z)$的概率密度为

$$p(x,y,z)=\frac{1}{(\sqrt{2\pi})^3\sigma^3}e^{-\frac{1}{2\sigma^2}(x^2+y^2+z^2)}$$

$S=\sqrt{x^2+y^2+z^2}$ 的分布函数为，当$s>0$时，

$$F(s)=P\{\sqrt{x^2+y^2+z^2}<s\}$$

$$=\iint\limits_{x^2+y^2+z^2<s^2}\frac{1}{(\sqrt{2\pi})^3\sigma^3}e^{-\frac{1}{2\sigma^2}(x^2+y+z^2)}dxdydz$$

作球面坐标变换，$x=\rho\cos\theta\sin\varphi, y=\rho\sin\theta\sin\varphi, z=\rho\cos\varphi$，则 $|J|=\rho^2\sin\varphi$，

$$F(s)=\int_0^{2\pi}d\theta\int_0^{\pi}\sin\varphi d\varphi\int_0^s\frac{1}{(\sqrt{2\pi})^3\sigma^3}e^{-\frac{1}{2}\rho^2/\sigma^2}\times$$

$$\times\rho^2 d\rho$$

$$=2\pi\cdot 2\int_0^s\frac{1}{(\sqrt{2\pi})^3\sigma^3}e^{-\frac{1}{2}\rho^2/\sigma^2}\cdot\rho^2 d\rho$$

由此式对s求导可得，当s>0时，S的密度函数为

$$F'(s)=f(s)=\sqrt{\frac{2}{\pi}}\frac{s^2}{\sigma^3}\exp\left(-\frac{s^2}{2\sigma^2}\right)$$

35．设ξ，η是两个独立随机变量，ξ服从N(0,1)，η服从自由度为n的χ^2分布(3.14)，令$t=\xi/\sqrt{\eta/n}$，试证t的密度函数为

$$P_n(x)=\frac{\Gamma(\frac{1}{2}(n+1))}{\sqrt{n\pi}\Gamma(\frac{1}{2}n)}\left(1+\frac{x^2}{n}\right)^{-\frac{1}{2}(n+1)}$$

这分布称为**具有自由度n的t－分布**，在数理统计中十分重要。

证：(3.14)式为

$$p(x)=\frac{1}{2^{\frac{1}{2}n}\Gamma(\frac{1}{2}n)}x^{\frac{1}{2}n-1}e^{-\frac{1}{2}x},\quad x>0.$$

令 $y=\sqrt{x/n}$，则$x=ny^2$，$x_y'=2ny$，由教本P146公式(10)
$p(y)=p[f^{-1}(y)]|[f^{-1}(y)]'|$得，

$\sqrt{\eta/n}$的密度函数为，当y>0时

$$p_{\sqrt{\eta/n}}(y)=\frac{(ny^2)^{\frac{1}{2}n-1}}{2^{\frac{1}{2}n}\Gamma(\frac{1}{2}n)}e^{-\frac{1}{2}ny^2}\cdot 2ny$$

$$=\frac{2n^{\frac{1}{2}n}y^{n-1}}{2^{\frac{1}{2}n}\Gamma(\frac{1}{2}n)}e^{-\frac{1}{2}ny^2}$$

ξ与$\sqrt{\eta/n}$仍独立。记$T=\xi/\sqrt{\eta/n}$，则由商的密度函数公式得T的密度函数为

$$p_T(t)=\int_{-\infty}^{\infty}|y|\,p_\xi(ty)\,p_{\sqrt{\eta/n}}(y)dy$$

$$=\int_0^{\infty}y\cdot\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}t^2y^2}\times$$

$$\times\frac{2n^{\frac{1}{2}n}\;y^{n-1}\;e^{-\frac{1}{2}ny^2}}{2^{\frac{1}{2}n}\Gamma(\frac{1}{2}n)}dy$$

$$=\int_0^{\infty}\frac{n^{\frac{1}{2}n}}{\sqrt{2\pi}\,2^{\frac{1}{2}n}\Gamma(\frac{1}{2}n)}\times$$

$$\times(y^2)^{\frac{1}{2}(n+1)-1}e^{-\frac{1}{2}y^2(n+t^2)}dy^2.$$

令$u=y^2(n+t^2)$，则$dy^2=du/(n+t^2)$，得

$$p_T(t)=\frac{n^{\frac{1}{2}n}(n+t^2)^{-\frac{1}{2}(n+1)}}{\sqrt{2\pi}\,2^{\frac{1}{2}n}\Gamma(\frac{1}{2}n)}\times$$

$$\times\int_0^{\infty}u^{\frac{1}{2}(n+1)-1}e^{-\frac{1}{2}u}du$$

$$=\frac{n^{\frac{1}{2}n}}{\sqrt{2\pi}\,2^{\frac{1}{2}n}\Gamma(\frac{1}{2}n)}\times$$

$$\times\frac{\Gamma(\frac{1}{2}(n+1))}{(1/2)^{\frac{1}{2}(n+1)}}(n+t^2)^{-\frac{1}{2}(n+1)}$$

$$\therefore\ p_T(t)=\frac{\Gamma(\frac{1}{2}(n+1))}{\sqrt{n\pi}\,\Gamma(\frac{1}{2}n)}\left(1+\frac{t^2}{n}\right)^{-\frac{1}{2}(n+1)}$$

$$-\infty<t<\infty$$

36. 设ξ,η,ζ有联合密度函数

$$f(x,y,z)=\begin{cases}6(1+x+y+z)^{-4}, & 当x>0,y>0,z>0时\\ 0, & 其它\end{cases}$$

试求$U=\xi+\eta+\zeta$的密度函数

解: U的分布函数为，当$t\leqslant 0$时$F(t)=0$；当$t>0$时有

$$F(t)=\iiint\limits_{x+y+z<t} p(x,y,z)dxdydz$$

$$=\int_0^t dx\int_0^{t-x} dy\int_0^{t-x-y}\frac{6}{(1+x+y)^4}dz$$

$$=\frac{-2}{(1+t)^3}\cdot\frac{t^2}{2}+\int_0^t dx\int_0^{t-x}\frac{2}{(1+x+y+z)^3}dy$$

$$=\frac{-t^2}{(1+t)^3}-\frac{t}{(1+t)^2}+\int_0^t\frac{1}{(1+x)^2}dx$$

$$=1-\frac{1}{t+1}-\frac{t}{(1+t)^2}-\frac{t^2}{(1+t)^3}$$

对F(t)求导可得U的密度函数为，当$t\leqslant 0$时 $p(t)=0$；当$t>0$时$p(t)=3t^2/(1+t)^4$。

37. 若ξ,η独立，且均服从$N(0,1)$，试证$U=\xi^2+\eta^2$与$V=\xi/\eta$是独立的。

证：(U,V)联合分布函数为

$$F(u,v)=\iint\limits_{\substack{x^2+y^2<u\\ x/y<v}}\frac{1}{2\pi}e^{-\frac{1}{2}(x^2+y^2)}dxdy.$$

当$s>0$时作变换，$s=x^2+y^2, t=x/y$，反函数有两支

$$\begin{cases}x=t\sqrt{s/(1+t^2)}\\ y=\sqrt{s/(1+t^2)}\end{cases}与\begin{cases}x=-t\sqrt{s/(1+t^2)}\\ y=-s\sqrt{s/(1+t^2)}\end{cases}$$

$$J^{-1}=\begin{vmatrix}2x & 2y\\ 1/y & -x/y^2\end{vmatrix}=-\frac{2x^2}{y^2}-2=-2(t^2+1),$$

$$|J|=\frac{1}{2(1+t^2)}$$

考虑到反函数有两支，分别利用两组变换得

$$F(u,v)=\left\{\iint\limits_{\substack{x^2+y^2<u\\ x/y<v,\,y>0}}+\iint\limits_{\substack{x^2+y^2<u\\ x/y<v,\,y<0}}\right\}\times$$

$$\times\frac{1}{2\pi}e^{-\frac{1}{2}(x^2+y^2)}dxdy$$

$$=2\int_0^u\int_{-\infty}^v\frac{1}{2\pi}e^{-\frac{1}{2}s}\cdot\frac{1}{2(1+t^2)}dt$$

对$F(u,v)$求导，得(U,V)的联合密度为（其余为0）

$$p(u,v)=\frac{1}{2}e^{-\frac{1}{2}u}\cdot\frac{1}{\pi(1+v^2)},\ u>0,\ 0<v<\infty$$

若令 $p_U(u)=\frac{1}{2}e^{-\frac{1}{2}u}\ (u>0)$，$p_V(v)=\frac{1}{\pi(1+v^2)}$

$(-\infty<v<\infty)$，

则U服从指数分布，V服从柯西分布，且$p(u,v)=p_U(u)\times p_V(v)$，所以U，V两随机变量独立。

38. 求证，如果ξ与η独立，且分别服从Γ－分布$G(\lambda,r_1)$和$G(\lambda,r_2)$，则$\xi+\eta$与ξ/η也独立。

证： 当$x>0$时，ξ与η的密度函数分别为

$$p_\xi(x)=\frac{\lambda^{r_1}}{\Gamma(r_1)}x^{r_1-1}e^{-\lambda x},$$

$$p_\eta(x)=\frac{\lambda^{r_2}}{\Gamma(r_2)}x^{r_2-1}e^{-\lambda x},$$

当$x\leqslant 0$时，$p_\xi(x)=p_\eta(x)=0$。设$U=\xi+\eta, V=\xi/\eta$。当$s\leqslant 0$或$t\leqslant 0$时，(U,V)联合密度为$p(s,t)=0$；当$s>0$，$t>0$时，作变换$s=x+y, t=x/y$，得$x=st/(1+t), y=s/(1+t)$，而$|J|=s/(1+t)^2$，所以

$$p(s,t)=\frac{\lambda^{r_1+r_2}}{\Gamma(r_1)\Gamma(r_2)}x^{r_1-1}y^{r_2-1}e^{-\lambda(x+y)}|J|$$

$$=\frac{\lambda^{r_1+r_2}}{\Gamma(r_1)\Gamma(r_2)}\left(\frac{st}{1+t}\right)^{r_1-1}\left(\frac{s}{1+t}\right)^{r_2-1}\times$$

$$\times e^{-\lambda s}\frac{s}{(1+t)^2}$$

$$=\left[\frac{\lambda^{r_1+r_2}}{\Gamma(r_1+r_2)}s^{r_1+r_2-1}e^{-\lambda s}\right]\times$$

$$\times\left[\frac{\Gamma(r_1+r_2)}{\Gamma(r_1)\Gamma(r_2)}\cdot\frac{t^{r_1-1}}{(1+t)^{r_1+r_2}}\right]$$

$$=p_U(s)p_V(t)。$$

由此知U服从Γ－分布$G(\lambda, r_1+r_2)$，V服从Beta分布，且U与V相互独立。

39．设独立随机变量ξ,η均服从

$$p(x)=\begin{cases}e^{-x}, & x>0\\ 0, & 其他,\end{cases}$$

问$\xi+\eta$与$\xi/(\xi+\eta)$是否独立？

解：令$U=\xi+\eta, V=\xi/(\xi+\eta)$。当$s\leqslant 0$或$t\overline{\in}(0,1)$时，U，V的联合密度$p(s,t)=0$；当$s>0$且$t\in(0,1)$时作变换$s=x+y, t=x/(x+y)$，则$x=st, y=s-st$，$|J|=s$，

$$p(s,t)=e^{-x}e^{-y}|J|=se^{-(x+y)}=se^{-s}\cdot 1=p_U(s)p_V(t)。$$

由此得U服从Γ－分布$G(1,2)$，V服从$(0,1)$均匀分布，且U

与V相互独立。

40.若(ξ,η)服从二元正态分布(2.22)，试找出ξ+η与ξ−η相互独立的充要条件。

解：(2.22)式为

$$p(x,y)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}}\exp\left\{-\frac{1}{2(1-r^2)}\times\right.$$

$$\left.\times\left[\frac{(x-a)^2}{\sigma_1^2}-\frac{2r(x-a)(y-b)}{\sigma_1\sigma_2}+\frac{(y-b)^2}{\sigma_2^2}\right]\right\}$$

设$U_1=\xi+\eta$，$V_1=\xi-\eta$；$U=U_1-a-b$，$V=V_1-a+b$。作变换$s=x+y-a-b$，$t=x-y-a+b$，则$x-a=\frac{1}{2}(s+t)$，$y-b=\frac{1}{2}(s-t)$，$|J|=\frac{1}{2}$。U，V的联合密度函数为

$$f(s,t)=p(x,y)\,|J|=\frac{1}{2}\cdot\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}}\times$$

$$\times\exp\left\{-\frac{1}{2(1-r^2)}\left[\frac{(s+t)^2}{4\sigma_1^2}-\right.\right.$$

$$\left.\left.-\frac{2r(s+t)(s-t)}{4\sigma_1\sigma_2}+\frac{(s-t)^2}{4\sigma_2^2}\right]\right\}$$

$$=\frac{1}{4\pi\sigma_1\sigma_2\sqrt{1-r^2}}\exp\left\{-\frac{1}{8(1-r^2)\sigma_1^2\sigma_2^2}\times\right.$$

$$\times\left[s^2(\sigma_1^2+\sigma_2^2-2r\sigma_1\sigma_2)+\right.$$

$$+t^2(\sigma_1^2+\sigma_2^2+2r\sigma_1\sigma_2)+$$

$$\left.\left.+2st\,(\sigma_2^2-\sigma_1^2)\right]\right\}$$

设U,V的边际分布密度函数分别为 $f_U(s)$,$f_V(t)$，欲U与V独立,必须且只需$f(s,t)=f_U(s)\cdot f_V(t)$,由$f(s,t)$的表达式可知，这当且仅当$\sigma_2^2-\sigma_1^2=0$时成立。U,V相互独立与$U_1$,$V_1$相互独立显然是等价的，所以$U_1=\xi+\eta$,$V_1=\xi-\eta$相互独立的充要条件是$\sigma_1=\sigma_2$。当$\sigma_1=\sigma_2=\sigma$时，得

$$f_U(s)=\frac{1}{2\sigma\sqrt{\pi(1+r)}}\exp\left\{-\frac{s^2}{4(1+r)\sigma^2}\right\},$$

$$f_V(t)=\frac{1}{2\sigma\sqrt{\pi(1-r)}}\exp\left\{-\frac{t^2}{4(1-r)\sigma^2}\right\}$$

$U\sim N(0,2(1+r)\sigma^2)$，$V\sim N(0,2(1-r)\sigma^2)$

41.对二元正态密度函数

$$p(x,y)=\frac{1}{2\pi}\exp\{-\tfrac{1}{2}(2x^2+y^2+2xy-22x-$$
$$-14y+65)\},$$

(1)把它化为标准形式(2.22)；(2)指出a,b,σ_1,σ_2,r；(3)求$p_1(x)$；(4)求$p(x|y)$。

解：(1)因为指数中二次项x^2,y^2,xy的系数分别为-1,$-\frac{1}{2}$,-1,所以与(2.22)式（见上题解答）比较知,可设其配方后的形式为

$$-1\cdot(x+s)^2-\tfrac{1}{2}(y+t)^2-1\cdot(x+s)(y+t).$$

比较系数得

$$\begin{cases}-2s-t=11\\ -s-t=7\\ -s^2-\frac{1}{2}t^2-st=32\frac{1}{2}\end{cases}$$

此方程组有唯一解$s=-4,t=-3$,由此得

$$p(x,y)=\frac{1}{2\pi}\exp\{-[(x-4)^2+\tfrac{1}{2}(y-3)^2+$$
$$+(x-4)(y-3)]\}$$

$$= \frac{1}{2\pi\cdot 1\cdot\sqrt{2}\sqrt{1-\frac{1}{2}}}\exp\left\{-\frac{1}{2(1-\frac{1}{2})}\times\right.$$

$$\times\left[(x-4)^2+\frac{(y-3)^2}{2}+2\times\right.$$

$$\left.\left.\times\frac{1}{\sqrt{2}}\cdot\frac{(x-4)(y-3)}{1\cdot\sqrt{2}}\right]\right\}$$

(2)与(2.22)式比较得，$a=4, b=3, \sigma_1=1, \sigma_2=\sqrt{2}$，$r=-1/\sqrt{2}$。

(3) $$p_1(x)=\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{(x-4)^2}{2}\right\},$$

$$p_2(y)=\frac{1}{2\sqrt{\pi}}\exp\left\{-\frac{(y-3)^2}{4}\right\}.$$

(4) $$p(x|y)=\frac{p(x,y)}{p_2(y)}=\frac{1}{\sqrt{\pi}}\exp\{-[x-(-\tfrac{1}{2}y+5\tfrac{1}{2})]^2\}$$
它服从$N(-\frac{1}{2}y+5\frac{1}{2}, \frac{1}{2})$。

42. 设$a=0$，$B^{-1}=\begin{pmatrix}7 & 3 & 2\\ 3 & 4 & 1\\ 2 & 1 & 2\end{pmatrix}$，试写出分布密度(2.12)，并求出$(\xi_1,\xi_2)$的边际密度函数。

解： $|B^{-1}|=27$，$|B|=1/|B^{-1}|=1/27$。

$$p(x,y,z)=\frac{1}{(2\pi)^{\frac{1}{2}n}|B|^{\frac{1}{2}}}\times$$

$$\times\exp\{-\tfrac{1}{2}(x-a)B^{-1}(x-a)^T\}$$

$$=\frac{1}{(2\pi)^{\frac{1}{2}n}|B|^{\frac{1}{2}}}\times$$

$$\times\exp\{-\frac{1}{2}\sum_{j,k=1}^{n}r_{jk}(x_j-a_j)(x_k-a_k)\}$$

$$= \frac{1}{(2\pi)^{3/2}\sqrt{1/27}} \times$$

$$\times \exp\{-\tfrac{1}{2}(7x^2+4y^2+2z^2+6xy+$$
$$+4xz+2yz)\}。$$

(ξ_1, ξ_2)的边际密度函数为（积分时在指数中对 z 配方）

$$p(x,y) = \int_{-\infty}^{\infty} p(x,y,z)dz$$

$$= \frac{1}{(2\pi)^{3/2}\sqrt{1/27}} e^{-\frac{1}{2}(5x^2+3\frac{1}{2}y^2+4xy)} \times$$

$$\times \int_{-\infty}^{\infty} e^{-(z+x+\frac{1}{2}y)^2} dz$$

令 $z+x+\frac{1}{2}y=t$，利用$\int_{-\infty}^{\infty} e^{-t^2} dt = \sqrt{\pi}$得

$$p(x,y) = \frac{3\sqrt{6}}{4\pi}\exp\{-\tfrac{1}{2}(5x^2+4xy+3\tfrac{1}{2}y^2)\}。$$

*43．设ξ，η是相互独立相同分布的随机变量，其密度函数不等于0，且有二阶导数，试证若$\xi+\eta$与$\xi-\eta$相互独立，则随机变量ξ，η，$\xi+\eta$，$\xi-\eta$均服从正态分布。

证：以 f 记 ξ 的密度函数，则(ξ, η)的联合密度为$f(x)f(y)$。作变换，令$s=x+y$，$t=x-y$，得$x=\frac{1}{2}(s+t)$，$y=\frac{1}{2}(s-t)$，$|J|=\frac{1}{2}$。若改记 s 为 x，t为y，则由此可得$(\xi+\eta, \xi-\eta)$的联合密度为$\frac{1}{2}f(\frac{1}{2}(x+y))f(\frac{1}{2}(x-y))$。另一方面，由卷积公式得$\xi+\eta$和$\xi-\eta$的密度分别为

$$g(x) = \int_{-\infty}^{\infty} f(x-s)f(s)ds,$$

$$h(y) = \int_{-\infty}^{\infty} f(y+t)f(t)dt,$$

故由$\xi+\eta$与$\xi-\eta$独立得

$$\frac{1}{2}f\left(\frac{1}{2}(x+y)\right)f\left(\frac{1}{2}(x-y)\right)=g(x)h(y).$$

令$m(x)=\log f(x)$（此处用了$f(x)>0$），则有

$$m\left(\frac{1}{2}(x+y)\right)+m\left(\frac{1}{2}(x-y)\right)=\log g(x)+\log 2h(y).$$

由假定知$m(x)$有二阶导数，上式对x求导得

$$m'\left(\frac{x+y}{2}\right)\left(\frac{x+y}{2}\right)'_x+m'\left(\frac{x-y}{2}\right)\left(\frac{x-y}{2}\right)'_x$$

$$=(\log g(x))'_x$$

再对y求一次导数得

$$\frac{1}{4}m''\left(\frac{1}{2}(x+y)\right)-\frac{1}{4}m''\left(\frac{1}{2}(x-y)\right)=0.$$

对任意u，v，选择x，y使$u=\frac{1}{2}(x+y)$，$v=\frac{1}{2}(x-y)$，则由上式得

$$m''(u)-m''(v)=0.$$

由u，v的任意性得$m''\equiv$常数，因而$m(x)=a+bx+cx^2$，即有

$$f(x)=\exp(a+bx+cx^2).$$

所以ξ，η，从而$\xi+\eta$，$\xi-\eta$均为正态分布。

*44．利用随机变量分布解释贝特朗奇论。

解：(1)将弦的一端A固定，另一端B在圆周上等可能分布，记ξ_1表示沿逆时针方向$\widehat{AB}$弧长，则ξ_1在$(0,2\pi)$上服从均匀分布，

$$P\{弦长>\sqrt{3}\}=P\left\{\frac{2\pi}{3}<\xi_1<\frac{4\pi}{3}\right\}=$$

$$=\int_{2\pi/3}^{4\pi/3}\frac{1}{2\pi}\,dx=\frac{1}{3}$$

(2)假定弦垂直于某直径，取该直径为x轴，圆心为坐标原点，记ξ_2表示弦的中点坐标，则ξ_2在$[-1,1]$上服从均匀分布，

$$P\{弦长>\sqrt{3}\}=P\left\{-\frac{1}{2}<\xi_2<\frac{1}{2}\right\}=$$

$$=\int_{-\frac{1}{2}}^{\frac{1}{2}}\frac{1}{2}\,dx=\frac{1}{2}$$

(3)以圆心为原点建立直角坐标系XOY，记弦中点的坐标为$\eta=(\eta_1,\eta_2)$，则η在圆内$\{(x,y):x^2+y^2\leqslant 1\}$服从均匀分布。记$D=\{(x,y):\ x^2+y^2<\frac{1}{2}\}$，则

$$P\{\text{弦长}>\sqrt{3}\}=P\{\eta\in D\}=$$

$$=\iint\limits_{x^2+y^2<\frac{1}{2}}\frac{1}{\pi}\,dxdy=\frac{1}{4}.$$

三种解法的随机变量虽都服从均匀分布，但由于随机变量不同，所以就得出了不同的结论。

*45．若f是Ω上单值实函数，对$B\subset R^1$，记 $f^{-1}(B)=\{\omega\in\Omega:\ f(\omega)\in B\}$。试证逆映照$f^{-1}$具有如下性质：

(1) $f^{-1}(\bigcup\limits_{\lambda\in\Lambda}B_\lambda)=\bigcup\limits_{\lambda\in\Lambda}f^{-1}(B_\lambda)$;

(2) $f^{-1}(\bigcap\limits_{\lambda\in\Lambda}B_\lambda)=\bigcap\limits_{\lambda\in\Lambda}f^{-1}(B_\lambda)$;

(3) $f^{-1}(\overline{B})=\overline{f^{-1}(B)}$.

证：(1)若$\omega\in f^{-1}(\bigcup\limits_{\lambda\in\Lambda}B_\lambda)$，则$f(\omega)\in\bigcup\limits_{\lambda\in\Lambda}B_\lambda$，必存在某个$\lambda_0\in\Lambda$使$f(\omega)\in B_{\lambda_0}$，亦有$\omega\in f^{-1}(B_{\lambda_0})$，从而

$\omega\in\bigcup\limits_{\lambda\in\Lambda}f^{-1}(B_\lambda)$,

$$\therefore\ \bigcup_{\lambda\in\Lambda}f^{-1}(B_\lambda)\supset f^{-1}(\bigcup_{\lambda\in\Lambda}B_\lambda).\tag{1}$$

反之，若$\omega\in\bigcup\limits_{\lambda\in\Lambda}f^{-1}(B_\lambda)$，必存在某个$\lambda_0\in\Lambda$使$\omega\in f^{-1}(B_{\lambda_0})$

亦有$f(\omega)\in B_{\lambda_0}$，即$f(\omega)\in\bigcup\limits_{\lambda\in\Lambda}B_\lambda$，从而$\omega\in f^{-1}(\bigcup\limits_{\lambda\in\Lambda}B_\lambda)$,

$$\therefore\quad f^{-1}(\bigcup_{\lambda\in\Lambda} B_\lambda)\supset\bigcup_{\lambda\in\Lambda} f^{-1}(B_\lambda). \qquad (2)$$

由(1),(2)式即得(和集的逆像等于每个集逆像的和)

$$f^{-1}(\bigcup_{\lambda\in\Lambda} B_\lambda)=\bigcup_{\lambda\in\Lambda} f^{-1}(B_\lambda).$$

(2)若$\omega\in f^{-1}(\bigcap_{\lambda\in\Lambda} B_\lambda)$,则$f(\omega)\in\bigcap_{\lambda\in\Lambda} B_\lambda$, 即$f(\omega)$属于每个$B_\lambda(\lambda\in\Lambda)$,得$\omega\in f^{-1}(B_\lambda)$(对任一$\lambda\in\Lambda$),从而

$\omega\in\bigcap_{\lambda\in\Lambda} f^{-1}(B_\lambda)$,

$$\therefore\quad \bigcap_{\lambda\in\Lambda} f^{-1}(B_\lambda)\supset f^{-1}(\bigcap_{\lambda\in\Lambda} B_\lambda). \qquad (3)$$

反之,若$\omega\in\bigcap_{\lambda\in\Lambda} f^{-1}(B_\lambda)$,则$\omega$属于每个$f^{-1}(B_\lambda)$ $(\lambda\in\Lambda)$,亦有$f(\omega)$属于每个$B_\lambda(\lambda\in\Lambda)$,即$f(\omega)\in\bigcap_{\lambda\in\Lambda} B_\lambda$,从而

$\omega\in f^{-1}(\bigcap_{\lambda\in\Lambda} B_\lambda)$,

$$\therefore\quad f^{-1}(\bigcap_{\lambda\in\Lambda} B_\lambda)\supset\bigcap_{\lambda\in\Lambda} f^{-1}(B_\lambda). \qquad (4)$$

由(3),(4)式得(交集的逆像等于每个集逆像的交)

$$f^{-1}(\bigcap_{\lambda\in\Lambda} B_\lambda)=\bigcap_{\lambda\in\Lambda} f^{-1}(B_\lambda).$$

(3)若$\omega\in f^{-1}(\overline{B})$,则$f(\omega)\overline{\in}B$,亦有$\omega\overline{\in}f^{-1}(B)$,

从而 $\omega\in\overline{f^{-1}(B)}$,所以$\overline{f^{-1}(B)}\supset f^{-1}(\overline{B})$。

反之,若$\omega\in\overline{f^{-1}(B)}$,则$\omega\overline{\in}f^{-1}(B)$,亦有$f(\omega)\overline{\in}B$,即$f(\omega)\in\overline{B}$,从而$\omega\in f^{-1}(\overline{B})$,所以$f^{-1}(\overline{B})\supset\overline{f^{-1}(B)}$。

由以上证明可得$f^{-1}(\overline{B})=\overline{f^{-1}(B)}$,即互为对立事件的

逆像也是互为对立的事件。

*46. 证明：ξ是一个随机变量，当且仅当对任何$x\in R^1$成立

$$\{\omega:\ \xi(\omega)<x\}\in F.$$

〔提示：必要性是显然的。为证充分性，记$M=\{A:A\subset R^1,(\xi(\omega)\in A)\in F\}$，验证M是$\sigma$－域，又M包含全体形如$(-\infty,x)$的区间，故M包含$B_1$〕。

证：必要性。设ξ是随机变量，则对$C\in B$有
$\{\omega:\ \xi(\omega)\in C\}\in F$，又$(-\infty,x)\in B_1$，

$$\therefore\ \{\omega:\ \xi(\omega)<x\}=\{\omega:\ \xi(\omega)\in(-\infty,x)\}\in F.$$

充分性。记$M=\{A:\ A\subset R^1,(\omega:\ \xi(\omega)\in A)\in F\}$，现证M是$R^1$中$\sigma$－域。

(1) $\{\omega:\ \xi(\omega)\in R^1\}=\Omega\in F$，故$R^1\in M$。

(2) 若$C\in M$，由上题$f^{-1}(\overline{C})=\overline{f^{-1}(C)}$得

$(\omega:\ \xi(\omega)\in\overline{C})=\Omega-(\omega:\ \xi(\omega)\in C)\in F$，故$\overline{C}\in M$，对余集运算封闭。

(3) 设$C_i\in M$，$i=1,2,\cdots$，由上题(1)中结论得$\bigcup\limits_{i=1}^{\infty}C_i\in M$，M关于可列并集运算封闭。

由(1)—(3)知，M是σ－域的集类。由条件知
$M\supset\{(-\infty,x):\ x\in R^1\}$，

$$\therefore\ M\supset S\{(-\infty,x):\ x\in R^1\}=B_1,$$

其中$S\{A\}$表示由集类A产生的σ－域。由此得证ξ是一随机变量。

第四章　数字特征与特征函数

本章内容　数学期望，方差，矩，熵与信息，母函数，特征函数，多元正态分布。

1. 设随机变量ξ只取非负整数值，其概率为$P\{\xi=k\}=a^k/(1+a)^{k+1}$，$a>0$是常数，试求$E\xi$及$D\xi$。

解. $E\xi=\sum\limits_{k=0}^{\infty}k\cdot\dfrac{a^k}{(1+a)^{k+1}}=\dfrac{1}{1+a}\left[\sum\limits_{k=1}^{\infty}k\left(\dfrac{a}{1+a}\right)^k\right]$，

令$a/(1+a)=p$，则$0<p<1$，且

$$\sum_{k=1}^{\infty}kp^k=p\left(\sum_{k=1}^{\infty}p^k\right)'=p\left(\frac{p}{1-p}\right)'=\frac{p}{(1-p)^2},$$

$$\therefore E\xi=\frac{1}{1+a}\cdot\frac{a}{1+a}\Big/\left(1-\frac{a}{1+a}\right)^2=a.$$

采用同样的方法并利用$E\xi=a$得

$$E\xi^2=\frac{1}{1+a}\left\{\sum_{k=1}^{\infty}k^2\left(\frac{a}{1+a}\right)^k\right\}$$

$$=\frac{1}{1+a}\sum_{k=1}^{\infty}k\left[(k-1)+1\right]p^k$$

$$=\frac{1}{1+a}\sum_{k=1}^{\infty}kp^k+\frac{1}{1+a}\sum_{k=1}^{\infty}k(k-1)p^k$$

$$= a + \frac{p^2}{1+a}\left(\sum_{k=1}^{\infty} p^k\right)'' = a + \frac{p^2}{1+a}\left(\frac{p}{(1-p)}\right)''$$

$$= a + \frac{p^2}{1+a} \cdot \frac{2}{(1-p)^3} = a + 2a^2$$

$D\xi = E\xi^2 - (E\xi)^2 = (a + 2a^2) - a^2 = a(1+a)$.

2.若事件A在第 i 次试验中出现的概率为p_i，设μ是事件A在起初 n 次独立试验中的出现次数，试求 Eμ 及 Dμ。

解. 设$\mu = \mu_1 + \mu_2 + \cdots + \mu_n$，其中

$$\mu_i = \begin{cases} 1, & \text{若第 } i \text{ 次试验A出现} \\ 0, & \text{若第 } i \text{ 次试验 } \bar{A} \text{ 出现,} \end{cases}$$

则 $E\mu = \sum_{i=1}^{n} E\mu_i = \sum_{i=1}^{n} p_i$.

由试验独立得诸 μ_i相互独立，由此得

$$D\mu = \sum_{i=1}^{n} D\mu_i = \sum_{i=1}^{n} p_i(1-p_i).$$

3.设 μ 是事件A在 n 次独立试验中的出现次数，在每次试验中P(A) = p，再设随机变量 η 视 μ 取偶数或奇数而取数值 0 及1，试求Eη及Dη。

解. η 服从两点分布，由第二章第 29 题得，P{η = 1} = P{事件A出现奇数次} = $\frac{1}{2} - \frac{1}{2}(1-2p)^n$，P{η = 0} = P{事件A出现偶数次} = $\frac{1}{2} + \frac{1}{2}(1-2p)^n$，所以

$$E\eta = \frac{1}{2} - \frac{1}{2}(1-2p)^n,$$

$$D\eta = \left[\frac{1}{2} - \frac{1}{2}(1-2p)^n\right]\left[\frac{1}{2} + \frac{1}{2}(1-2p)^n\right]$$

$$= \frac{1}{4} - \frac{1}{4}(1-2p)^{2n}.$$

4.袋中有 k 号的球 k 只，k = 1,2,…,n，从中摸出一球，求所得号码的数学期望。

解．设 ξ 表取一球的号码数．袋中球的总数为 $1+2+\cdots+n=\frac{1}{2}n(n+1)$，所以

$$P\{\xi=k\}=\frac{k}{\frac{1}{2}n(n+1)}$$

$$=\frac{2k}{n(n+1)},k=1,2,\cdots,n.$$

$$E\xi=\sum_{k=1}^{n}\frac{2k}{n(n+1)}\cdot k$$

$$=\frac{2}{n(n+1)}\cdot\frac{n(n+1)(2n+1)}{6}$$

$$=\frac{1}{3}(2n+1).$$

5．随机变量 μ 取非负整数值 $n\geqslant 0$ 的概率为 $p_n=AB^n/n!$，已知 $E\mu=a$，试决定A与B。

解．由于 μ 是分布，所以应有

$$\sum_{n=0}^{\infty}P\{\mu=n\}=\sum_{n=0}^{\infty}A\cdot\frac{B^n}{n!}=1,\text{即}Ae^B=1,A=e^{-B}.$$

又由已知

$$E\mu=\sum_{n=0}^{\infty}n\cdot\frac{AB^n}{n!}=a,\text{即}AB\sum_{n=1}^{\infty}\frac{B^{n-1}}{(n-1)!}=a,$$

$$ABe^B=a,\quad\therefore B=a,\quad A=e^{-B}=e^{-a}.$$

6．一袋中含有 a 只白球，b只黑球，从中摸出 c 个（$c\leqslant a+b$），求摸出白球数 μ 的数学期望。

解．μ 表示摸出 c 个球中白球个数，摸 c 个球可视为不放回地摸 c 次．记

$$\xi_i=\begin{cases}1, & \text{第 } i \text{ 次摸到白球}\\ 0, & \text{第 } i \text{ 次摸到黑球,}\end{cases}$$

则$\mu=\xi_1+\xi_2+\cdots\xi_c$。由第二章第7题得$P\{\xi_i=1\}=$
$=a/(a+b)$，$i=1, 2, \cdots, c$。所以 $E\xi_i=a/(a+b)$，

$$E\mu=\sum_{i=1}^{c}E\xi_i=c\cdot\frac{a}{a+b}=\frac{ac}{a+b}.$$

7．袋中有 n 张卡片，记号码1，2…，n，从中有放回地抽出 k 张卡片来，求所得号码之和 μ 的数学期望及方差。

解．记 μ 表示抽出 k 张卡片的号码和，ξ_i 表示第 i 次抽到卡片的号码，则$\mu=\xi_1+\xi_2+\cdots+\xi_k$。因为是放回抽取，所以诸 ξ_i 独立。由此得，对 $i=1, 2, \cdots, k$，

$$E\xi_i=\sum_{j=1}^{n}j\cdot\frac{1}{n}=\frac{1}{n}\sum_{j=1}^{n}j$$

$$=\frac{1}{n}\cdot\frac{n(n+1)}{2}=\frac{n+1}{2},$$

$$E\mu=E\xi_1+E\xi_2+\cdots+E\xi_k=\tfrac{1}{2}k(n+1);$$

$$E\xi_i^2=\sum_{j=1}^{n}j^2\cdot\frac{1}{n}=\frac{1}{n}\cdot\frac{n(n+1)(2n+1)}{6}$$

$$=\tfrac{1}{6}(n+1)(2n+1),$$

$$D\xi_i=E\xi_i^2-(E\xi_i)^2$$

$$=\tfrac{1}{6}(n+1)(2n+1)-\tfrac{1}{4}(n+1)^2$$

$$=\frac{1}{12}(n^2-1),$$

$$D\mu=D\xi_1+D\xi_2+\cdots+D\xi_n=\frac{1}{12}k(n^2-1)$$

8．在上题中，若采用不放回摸卡片，求所得号码之和的数学期望与方差。

解．设μ为所得k张卡片上号码之和。对$1\leqslant i_1<\cdots<i_k\leqslant n$

有 $P\{\mu=i_1+i_2+\cdots+i_k\}=1/C_n^k$，由定义得

$$E\mu=\sum_{1\leqslant i_1<\cdots<i_k\leqslant n}(i_1+i_2+\cdots+i_k)\frac{1}{C_n^k}$$

$$=\frac{C_{n-1}^{k-1}}{C_n^k}\sum_{m=1}^{n}m \qquad (*)$$

每次抽卡片 k 张称为一组，对于每个固定的卡片m，在卡片m所在的组中，其余k－1张卡片可以从剩下n－1张卡片中任意抽取，所以 m 总共被抽到的次数（或所在的组数）为 C_{n-1}^{k-1}，转换成对 m 求和就得到上式。由此得

$$E\mu=\frac{k}{n}\cdot\frac{n(n+1)}{2}=\frac{k(n+1)}{2}.$$

为求方差，先求$E\mu^2$，由定义得

$$E\mu^2=\sum_{1\leqslant i_1<\cdots<i_k\leqslant n}(i_1+\cdots+i_k)^2\cdot\frac{1}{C_n^k}$$

$$=\frac{1}{C_n^k}\sum_{1\leqslant i_1<\cdots<i_k\leqslant n}\Big\{(i_1^2+\cdots+i_k^2)+$$

$$+2\sum_{1\leqslant t<j\leqslant k}i_ti_j\Big\}$$

$$=\frac{C_{n-1}^{k-1}}{C_n^k}\sum_{m=1}^{n}m^2+2\frac{C_{n-2}^{k-2}}{C_n^k}\sum_{1\leqslant i<j\leqslant n}i\,j$$

其中前一个和式得来的理由同（*）式。为获得后一和式，仍考虑每次抽得的一组 k 张卡片。在卡片 i 和 j 所在的组中，

其余k-2张卡片可以从其余n-2张卡片中任意抽取，所以卡片 i 和 j 同时被抽到的次数为 C_{n-2}^{k-2}，即得第二个和式的系数。继续运算得

$$E\mu^2=\frac{k}{n}\sum_{m=1}^{n}m^2+\frac{2\,C_{n-2}^{k-2}}{C_n^k}\left[1(2+\cdots+n)+2(3+\cdots+n)+\cdots+(n-1)n\right]$$

$$=\frac{k}{n}\cdot\frac{n(n+1)(2n+1)}{6}+\frac{2k(k-1)}{n(n-1)}\sum_{m=1}^{n-1}m\left[\frac{n(n+1)}{2}-\frac{m(m+1)}{2}\right]$$

$$=\frac{k(n+1)(2n+1)}{6}+\frac{2k(k-1)}{n(n-1)}\times\left[\frac{n(n+1)}{2}\cdot\frac{n(n-1)}{2}-\frac{1}{2}\sum_{m=1}^{n-1}m^3-\frac{1}{2}\sum_{m=1}^{n-1}m^2\right]$$

$$=\frac{k(n+1)(2n+1)}{6}+k(k-1)\left[\frac{n(n+1)}{2}-\frac{(n-1)^2n^2}{4n(n-1)}-\frac{(n-1)n(2n-1)}{6n(n-1)}\right]$$

$$=\frac{1}{6}k(n+1)(2n+1)+k(k-1)\left[\frac{1}{2}n(n+1)-\frac{1}{4}n(n-1)-\frac{1}{6}(2n-1)\right]$$

$$=\frac{1}{6}k(n+1)(2n+1)+\frac{1}{12}k(k-1)(n+1)(3n+2),$$

$$\therefore\quad D\mu=E\mu^2-(E\mu)^2=\frac{1}{12}k(n+1)(n-k).$$

9.试证：若取非负整数值的随机变量 ξ 的数学期望存在，则

$$E\xi=\sum_{k=1}^{\infty}P\{\xi\geqslant k\}.$$

证：$\sum_{k=1}^{\infty}P\{\xi\geqslant k\}=\sum_{k=1}^{\infty}\quad\sum_{j=k}^{\infty}P\{\xi=j\}$

$$\begin{aligned}&=\{\xi=1\}+P\{\xi=2\}+P\{\xi=3\}+\cdots\\&\qquad+P\{\xi=2\}+P\{\xi=3\}+\cdots\\&\qquad\qquad+P\{\xi=3\}+\cdots\\&=\sum_{k=1}^{\infty}kP\{\xi=k\}=E\xi.\end{aligned}$$

10.某城市共有 N 辆汽车，车牌号从 1 到 N，若随机地记下 n 辆车的车牌号，其最大号码为 ξ，求 Eξ.

解：此题属于有放回抽样的情形，利用允许重复的排列计算.若 n 辆车的车牌号中最大号码为 k，则其中应至少有一个牌号为 k，所以有利场合数应为，k^n与那些 n 个牌号中没有号码 k 的种数$(k-1)^n$之差，从而有

$$P\{\xi=k\}=[k^n-(k-1)^n]/N^n,\ k=1,2,\cdots,N.$$

$$\therefore\quad E\xi=\sum_{k=1}^{N}k\cdot\frac{k^n-(k-1)^n}{N^n}=N-\sum_{k=1}^{N-1}\frac{k^n}{N^n}.$$

11.若随机变量 ξ 服从拉普拉斯分布，其密度函数为

$$p(x)=\frac{1}{2\lambda}e^{-|x-\mu|/\lambda},\quad -\infty<x<\infty,\ \lambda>0.$$

试求 Eξ，Dξ.

解：$E\xi=\int_{-\infty}^{\infty}\frac{1}{2\lambda}xe^{-|x-\mu|/\lambda}dx$ （令$t=(x-\mu)/\lambda$）

$$=\int_{-\infty}^{\infty}\frac{\lambda t+\mu}{2}e^{-|t|}dt$$

$$=\int_{-\infty}^{\infty}\frac{\lambda t}{2}\ e^{-|t|}dt+\int_{-\infty}^{\infty}\frac{\mu}{2}\ e^{-|t|}dt$$

$$=0+\mu=\mu。$$

$$D\xi=\int_{-\infty}^{\infty}\frac{1}{2\lambda}(x-\mu)^2e^{-|x-\mu|/\lambda}\quad(令t=(x-\mu)/\lambda)$$

$$=\lambda^2\int_0^{\infty}t^2e^{-t}dt$$

$$=\lambda^2t^2(-e^{-t})\Big|_0^{\infty}+2\lambda^2\int_0^{\infty}te^{-t}\,dt$$

$$=2\lambda^2t(-e^{-t})\Big|_0^{\infty}+2\lambda^2\int_0^{\infty}e^{-t}dt$$

$$=2\lambda^2\ (-e^{-t})\Big|_0^{\infty}=2\lambda^2。$$

12.若分子的速度的分布密度函数由马克斯威尔分布律给出

$$p(x)=\begin{cases}\dfrac{4x^2}{\alpha^3\sqrt{\pi}}e^{-x^2/\alpha^2}, & x>0\\ 0, & x\leqslant 0\end{cases}$$

其中$\alpha>0$是常数，试求分布的平均速度和平均动能（假定分子的质量等于m）。

解：分子平均速度为

$$E\xi=\int_0^{\infty}x\frac{4x^2}{\alpha^3\sqrt{\pi}}e^{-x^2/\alpha^2}dx(令t=x/\alpha)$$

$$=\int_0^{\infty}\frac{4t^3}{\sqrt{\pi}}e^{-t^2}\cdot\alpha dt$$

$$= \frac{2\alpha t^2}{\sqrt{\pi}}(-e^{-t^2})\Big|_0^{\infty} + \frac{2\alpha}{\sqrt{\pi}}\int_0^{\infty} 2te^{-t^2}dt = \frac{2\alpha}{\sqrt{\pi}}。$$

分子平均动能为

$$E(\tfrac{1}{2}m\xi^2) = \int_0^{\infty}\tfrac{1}{2}mx^2\frac{4x^2}{\alpha^3\sqrt{\pi}}e^{-x^2/\alpha^2}dx$$

$$(令\ t = \frac{x}{\alpha})$$

$$= \int_0^{\infty}\frac{2m\alpha^2}{\sqrt{\pi}}t^4e^{-t^2}dt$$

$$= \frac{m\alpha^2}{\sqrt{\pi}}t^3(-e^{-t^2})\Big|_0^{\infty} +$$

$$+ \frac{3m\alpha^2}{\sqrt{\pi}}\int_0^{\infty}t^2e^{-t^2}dt$$

$$= \frac{3m\alpha^2}{2\sqrt{\pi}}t(-e^{-t^2})\Big|_0^{\infty} + \frac{3m\alpha^2}{2\sqrt{\pi}}\int_0^{\infty}e^{-t^2}dt$$

$$= \frac{3m\alpha^2}{2\sqrt{\pi}} \times \frac{\sqrt{\pi}}{2} = \frac{3m\alpha^2}{4}$$

13. 若ξ_1，ξ_2相互独立，均服从$N(a,\sigma^2)$，试证

$$E\max(\xi_1, \xi_2) = a + \sigma/\sqrt{\pi}。$$

证：$\xi_1\xi_2$ 的联合密度为

$$p(x,y) = \exp\left\{-\frac{(x-a)^2}{2\sigma^2} - \frac{(y-a)^2}{2\sigma^2}\right\},$$

$$\therefore\quad E\max(\xi_1,\xi_2) = \iint \max(x,y)p(x,y)dxdy$$

$$= \int_{-\infty}^{\infty}dx\int_{-\infty}^{x}xp(x,y)dy +$$

$$+ \int_{-\infty}^{\infty}dx\int_{x}^{\infty}yp(x,y)dy$$

（利用密度函数的积分值为1，减 a 再加 a）

$$=\int_{-\infty}^{\infty} dx \int_{-\infty}^{x} (x-a)d(x, y)dy +$$

$$+\int_{-\infty}^{\infty} dx \int_{x}^{\infty}(y-a)p(x,y)dy + a$$

(在前一积分中交换积分次序，在后一积分中交换 x 与y 的记号)

$$=\int_{-\infty}^{\infty} dy \int_{y}^{\infty}(x-a)p(x,y)dx +$$

$$+\int_{-\infty}^{\infty} dy \int_{y}^{\infty} (x-a)p(y,x)dx + a$$

$$=a+2\int_{-\infty}^{\infty}\frac{1}{2\pi\sigma^2} e^{-\frac{(y-a)^2}{2\sigma^2}}dy \times$$

$$\times\int_{y}^{\infty} (x-a)e^{-\frac{(x-a)^2}{2\sigma^2}}dx$$

$$=a+\frac{1}{\pi}\int_{-\infty}^{\infty} e^{-(y-a)^2/\sigma^2} dy$$

$$\left(令 \quad (y-a)/\sigma = t\right)$$

$$= a+\frac{1}{\pi}\int_{-\infty}^{\infty} \sigma e^{-t^2} dt$$

$$=a+\frac{\sigma}{\pi}\sqrt{\pi}= a+\sqrt{\frac{\sigma}{\pi}}。$$

14.若随机变量 ξ 的分布函数为 F(x)，试证

$$E\xi=\int_{0}^{\infty}\left(1-F(x)\right) dx-\int_{-\infty}^{0} F(x)dx.$$

证： $E\xi=\int_{-\infty}^{\infty} xdF(x)=\int_{-\infty}^{0} xdF(x)+\int_{0}^{\infty}xdF(x)$

$$=\int_{-\infty}^{0} xdF(x)-\int_{0}^{\infty} xd\left(1-F(x)\right)$$

$$=xF(x)\Big|_{-\infty}^{0}-\int_{-\infty}^{0}F(x)\,dx-$$

$$-x\Big(1-F(x)\Big)\Big|_{0}^{\infty}+\int_{0}^{\infty}\Big(1-F(x)\Big)dx$$

由均值存在得$\int_{-\infty}^{\infty}|x|dF(x)<\infty$,

$$\therefore\quad 0\leqslant AF(-A)\leqslant\int_{-\infty}^{-A}|x|\,dF(x)\to 0\ (\text{当}A\to+\infty),$$

$$0\leqslant B(1-F(B))\leqslant\int_{B}^{\infty}|x|\,dF(x)\to 0(\text{当}B\to+\infty),$$

以此代入 $E\xi$ 的计算式即得

$$E\xi=\int_{0}^{\infty}(1-F(x))\,dx-\int_{-\infty}^{0}F(x)dx.$$

15.若 ξ_1, ξ_2, …, ξ_n 为正的独立随机变量，服从相同分布，密度函数为 $f(x)$，试证

$$E\left(\frac{\xi_1+\xi_2+\cdots+\xi_k}{\xi_1+\xi_2+\cdots+\xi_n}\right)=\frac{k}{n}$$

证：由$\xi_i>0$知 $\dfrac{\xi_1(\omega)+\cdots+\xi_k(\omega)}{\xi_1(\omega)+\cdots+\xi_n(\omega)}$ 分母不为0，它是有限可测函数，即是随机变量，利用独立同分布可得

$$E\left(\frac{\xi_1+\cdots+\xi_k}{\xi_1+\cdots+\xi_n}\right)=\int\cdots\int\frac{x_1+\cdots+x_k}{x_1+\cdots+x_n}\times$$

$$\times f(x_1)\cdots f(x_n)dx_1\cdots dx_n$$

$$=k\int\cdots\int\frac{x_1}{x_1+\cdots+x_n}f(x_1)\cdots f(x_n)dx_1\cdots dx_n$$

$$=\frac{k}{n}\int\cdots\int\frac{x_1+\cdots+x_n}{x_1+\cdots+x_n}f(x_1)\cdots f(x_n)dx_1\cdots dx_n$$

$$=\frac{k}{n}.$$

16. 袋中装有 N 只球，但其中白球数为随机变量，只知其数学期望为 n，试求从该袋中摸一球得到白球的概率。

解：记 ξ 为袋中的白球数，则由题设得

$$n=\sum_{k=1}^{N}kP\{\xi=k\}$$

由此利用全概率公式得

$$P\{\text{摸一球为白球}\}=$$

$$=\sum_{N>k>0,P(\xi=k)>0}P\{\text{摸一球为白球}\mid\xi=k\}P\{\xi=k\}$$

$$=\sum_{k=0}^{N}P\{\xi=k\}\cdot\frac{k}{N}=\frac{1}{N}\sum_{k=1}^{N}kP\{\xi=k\}=\frac{n}{N}$$

17. 甲袋中有 a 只白球 b 只黑球，乙袋中装有 α 只白球 β 只黑球，现从甲袋中摸出 $c(c\leqslant a+b)$ 只球放入乙袋中，求从乙袋中再摸一球而为白球的概率。

解：令 B 表"从乙袋摸一球为白球"，ξ 表从甲袋所摸 c 个球中白球数，则 ξ 取值 0，1，…，c，服从超几何分布，且 $E\xi=ca/(a+b)$（见教本P173）。考虑到若 $c>a$，则当 $i=a+1$，…，c 时 $P\{\xi=i\}=0$；若 $c>b$，则当 $i<c-b$ 时 $P\{\xi=i\}=0$；而在条件概率定义中要求 $P(A_i)=P\{\xi=i\}>0$，由此得

$$P(B)=\sum_{i=\max(0,c-b)}^{\min(a,c)}P\{\xi=i\}P\{B\mid\xi=i\}$$

$$=\sum_{i=0}^{c}P\{\xi=i\}\frac{\alpha+i}{\alpha+\beta+c}$$

$$=\frac{\alpha}{\alpha+\beta+c}\sum_{i=0}^{c}P\{\xi=i\}+$$

$$+\frac{1}{\alpha+\beta+c}\sum_{i=0}^{a} iP\{\xi=i\}$$

$$=\frac{\alpha}{\alpha+\beta+c}+\frac{E\xi}{\alpha+\beta+c}$$

$$=\frac{1}{\alpha+\beta+c}\left(\alpha+\frac{ac}{a+b}\right).$$

18.袋中有 a 只白球 b 只黑球，每次摸出一球后总是放入一只白球，这样进行了 n 次后，再从袋中摸一只球，求它是白球的概率。

解法一：设 ξ_i，η_i 分别表示进行 i 次后袋中白，黑球数，$p_i^{(1)}$，$p_i^{(2)}$分别表示第 i 次从袋中摸一球是白，黑球的概率，则由本章第16题的结论得

$$\left.\begin{aligned}&E\xi_i+E\eta_i=a+b,\\&p_{i+1}^{(1)}=E\xi_i/(a+b),\ p_{i+1}^{(2)}=E\eta_i/(a+b).\end{aligned}\right\}\qquad(1)$$

$$令\zeta_i=\begin{cases}1，第 i 次摸出一个黑球\\0，第 i 次摸出一个白球，\end{cases}$$

则$E\zeta_i=p_i^{(2)}$，$i=1, 2, \cdots$。

从袋中摸出一个黑球的概率是 $p_1^{(2)}$，也就是换一个黑球为白球的概率是 $p_1^{(2)}$，袋中原有 a 个白球，所以这时袋中白球数期望为 $E\xi_1=a+E\zeta_1=a+p_1^{(2)}$。接着再摸第二次，摸出一个黑球，也就是换一个黑球为白球的概率是 $p_2^{(2)}$，这时袋中白球数期望为 $E\xi_2=a+p_1^{(2)}+p_2^{(2)}$，等等。由此推得一般，并利用(1)中三式变形得

$$E\xi_n=a+\sum_{i=1}^{n}p_i^{(2)}=a+\sum_{i=1}^{n}\frac{E\eta_{i-1}}{a+b}$$

$$=a+\frac{1}{a+b}\sum_{i=1}^{n}(a+b-E\xi_{i-1}). \qquad (2)$$

置上式中 n 为 n－1 得

$$E\xi_{n-1}=a+\frac{1}{a+b}\sum_{i=1}^{n-1}(a+b-E\xi_{i-1}). \qquad (3)$$

(2)－(3)得

$$E\xi_n-E\xi_{n-1}=\frac{1}{a+b}(a+b-E\xi_{n-1})$$

$$=1-\frac{1}{a+b}E\xi_{n-1}.$$

由此式可得递推公式，依次递推并利用等比级数求和公式得（初始条件$E\xi_0=a$）：

$$E\xi_n=1+\frac{a+b-1}{a+b}E\xi_{n-1}=$$

$$=1+\frac{a+b-1}{a+b}+\left(\frac{a+b-1}{a+b}\right)^2E\xi_{n-2}=\cdots$$

$$=1+\left(\frac{a+b-1}{a+b}\right)+\left(\frac{a+b-1}{a+b}\right)^2+\cdots$$

$$+\left(\frac{a+b-1}{a+b}\right)^{n-1}+a\left(\frac{a+b-1}{a+b}\right)^n$$

$$=\left(\left(\frac{a+b-1}{a+b}\right)^n-1\right)\div\left(\frac{a+b-1}{a+b}-1\right)+$$

$$+a\left(\frac{a+b-1}{a+b}\right)^n$$

$$=(a+b)-\left(\frac{a+b-1}{a+b}\right)^n(a+b-a).$$

所以进行 n 次后，再从袋中摸一球是白球的概率为

$$p_{n+1}^{(1)}=\frac{E\xi_n}{a+b}=1-\frac{b}{a+b}\left(\frac{a+b-1}{a+b}\right)^n.$$

解法二：以A表事件“进行 i 次后，再从袋中摸一球是白球”，$i=1,2,\cdots,n$。则由全概率公式得

$$P(A_{i+1})=P(A_i)P(A_{i+1}|A_i)+P(\bar{A}_i)P(\bar{A}_{i+1}|\bar{A}_i)$$
$$=P^2(A_i)+[1-P(A_i)]\left(P(A_i)+\frac{1}{a+b}\right)$$
$$=P(A_i)\left(\frac{a+b-1}{a+b}\right)+\frac{1}{a+b}.$$

利用此递推公式即可求得欲求的概率。

19. 甲袋中有 a 只白球 b 只黑球，乙袋中有 c 只白球 d 只黑球，从两袋中各摸出一球，并交换放入另一袋中，这样做了 n 次之后，再从甲袋中摸出一球，求这球是白球的概率。

解：以ξ_i，η_i 分别表示交换 i 次后甲、乙两袋中白球数，第 $i+1$ 次由甲，乙两袋中取出一白球的概率分别为 $p_{i+1}^{(1)}$，$p_{i+1}^{(2)}$，则由本章第16题的结论得

$$\left.\begin{aligned}&E\xi_i+E\eta_i=a+c,\ p_{i+1}^{(1)}=E\xi_i/(a+b),\\&\qquad\qquad p_{i+1}^{(2)}=E\eta_i/(c+d).\end{aligned}\right\}\quad(*)$$

令$\zeta_i^{(1)}\left(\zeta_i^{(2)}\right)=\begin{cases}1\ ,\ \text{第 } i \text{ 次从甲(乙)袋摸出一个白球}\\0\ ,\ \text{第 } i \text{ 次从甲(乙)袋摸出一个黑球,}\end{cases}$

显然有 $E\xi_1=a-E\zeta_1^{(1)}+E\zeta_1^{(2)}=a-p_1^{(1)}+p_1^{(2)}$,

$$E\xi_2=a-E\zeta_1^{(1)}+E\zeta_1^{(2)}-E\zeta_2^{(1)}+E\zeta_2^{(2)}$$
$$=a-p_1^{(1)}-p_2^{(1)}+p_1^{(2)}+p_2^{(2)},$$

由此可推得一般，并利用(*)中三式变形得

$$E\xi_n = a - \sum_{i=1}^{n} p_i^{(1)} + \sum_{i=1}^{n} p_i^{(2)}$$

$$= a + \sum_{i=1}^{n}\left(\frac{E\eta_{i-1}}{c+d} - \frac{E\xi_{i-1}}{a+b}\right)$$

$$= a + \sum_{i=1}^{n}\left(\frac{a+c-E\xi_{i-1}}{c+d} - \frac{E\xi_{i-1}}{a+b}\right).$$

置上式中 n 为 n-1 可得$E\xi_{n-1}$的表达式，以$E\xi_n$减 $E\xi_{n-1}$并整理可得递推关系式

$$E\xi_n = \frac{a+c}{c+d} + E\xi_{n-1}\left(1 - \frac{1}{a+b} - \frac{1}{c+d}\right).$$

由于 $E\xi_0 = a$，由上式递推并利用等比级数求和公式可得

$$E\xi_n = \frac{(a+c)(a+b)}{a+b+c+d} + \frac{a(c+d)-c(a+b)}{a+b+c+d}\left(1 - \frac{1}{a+b} - \frac{1}{c+d}\right)^n.$$

所以题中欲求的概率为

$$p_{n+1}^{(1)} = \frac{E\xi_n}{a+b} = \frac{a+c}{a+b+c+d} + \frac{ad-cb}{(a+b)(a+b+c+d)}\left(1 - \frac{1}{a+b} - \frac{1}{c+d}\right)^n.$$

20．现有 n 个袋子，各装有 a 只白球 b 只黑球，先从第一个袋子中摸出一球，记下颜色后就把它放入第二个袋子中，再从第二个袋子中摸出一球，记下颜色后就把它放入第三个袋子中，照这样办法依次摸下去，最后从第 n 个袋中摸出一球并记下颜色，若在这 n 次摸球中所摸得的白球总数为 S_n，求ES_n。

解：令 $\xi_i=\begin{cases}1，从第 i 个袋子中摸出白球\\0，从第 i 个袋子中摸出黑球，\end{cases}$ 则

$$P\{\xi_1=1\}=a/(a+b),$$

$$P\{\xi_2=1\}=\frac{a}{a+b}\cdot\frac{a+1}{a+b+1}+$$

$$+\frac{b}{a+b}\cdot\frac{a}{a+b+1}$$

$$=\frac{a^2+ab+a}{(a+b)(a+b+1)}=\frac{a}{a+b}.$$

由此类推得 $P\{\xi_i=1\}=a/(a+b)$，$i+1, 2, \cdots, n$。又 $S_n=\xi_1+\xi_2+\cdots+\xi_n$，

$$\therefore\quad ES_n=\sum_{i=1}^{n}E\xi_i=\frac{na}{a+b}.$$

21.在物理实验中，为测量某物体的重量，通常要重复测量多次，最后再把测量记录的平均值作为该物体的重量，试说明这样做的道理。

解：以 ξ_i 表第 i 次测量值，由于受测量过程中许多随机因素的影响，测量值 ξ_i 和物体真实重量 a 之间有偏差，ξ_i 是独立同分布的随机变量，并有 $E\xi_i=a$，$D\xi_i=\sigma^2$。测量记录的平均值记为 η，则 $\eta=\frac{1}{n}(\xi_1+\cdots+\xi_n)$

$$E\eta=\frac{1}{n}\sum_{i=1}^{n}E\xi_i=\frac{na}{n}=a,$$

$$D\eta=\frac{1}{n^2}\sum_{i=1}^{n}D\xi_i=\frac{n\sigma^2}{n^2}=\frac{\sigma^2}{n}.$$

平均值 η 的均值仍为 a，但方差只有 ξ_i 方差的 $^1/_n$，而方

差是描述随机变量对于其数学期望的离散程度，所以以 η 作为物体的重量，则更近于真值。

22. 若 $\xi_1,\xi_2,\cdots,\xi_n$ 是独立随机变量，$D\xi_i=\sigma_i^2$，试找"权" $a_1,a_2,\cdots,a_n$（它们满足 $\sum_{i=1}^n a_i=1$），使 $\sum_{i=1}^n a_i\xi_i$ 的方差最小。

解. 若有某 $\sigma_k=0$，取 $a_k=1$，其余 $a_i=0$，其方差最小。若所有 $\sigma_i\neq 0$，则可视方差为 n 个变元 $a_1,a_2,\cdots,a_n$ 的函数，

$$D\left(\sum_{i=1}^n a_i\xi_i\right)=\sum_{i=1}^n a_i^2\sigma_i^2=f(a_1,\ a_2,\ \cdots,\ a_n),$$

由题设知有联系方程

$$\Phi(a_1,a_2,\cdots a_n)=\sum_{i=1}^n a_i-1=0\ .$$

现用拉格朗日乘数法求方差何时最小，为此作函数

$$F(a_1,a_2,\cdots,a_n)=f+\lambda\Phi$$

$$=\sum_{i=1}^n a_i^2\sigma_i^2+\lambda\left(\sum_{i=1}^n a_i-1\right)$$

令 $\dfrac{\partial F}{\partial a_i}=2a_i\sigma_i^2+\lambda=0$，$i=1,2,\cdots,n$

解得 $a_i=-\dfrac{\lambda}{2\sigma_i^2}$，又 $\sum_{i=1}^n\dfrac{-\lambda}{2\sigma_i^2}=1$，

从而 $\lambda=-1\Big/\sum_{i=1}^n\dfrac{1}{2\sigma_i^2}$，

所以当$a_i = \dfrac{1}{\sigma_i^2 \sum\limits_{k=1}^{n} \dfrac{1}{\sigma_k^2}}$时方差最小，这时方差为

$$D\left(\sum_{i=1}^{n} \xi_i\right) = \sum_{i=1}^{n} a_i^2 \sigma_i^2 = 1\Big/\sum_{i=1}^{n} \frac{1}{\sigma_i^2}.$$

若σ_i相等，则 $a_i = \dfrac{1}{n}$，$D\left(\sum\limits_{i=1}^{n} \xi_i\right) = \dfrac{1}{n}\sigma^2$。

23．设随机变量ξ_1，ξ_2，…，$\xi_{m+n}(n>m)$是独立的，有相同的分布并且有有限的方差，试求$S=\xi_1+\cdots+\xi_n$与$T=\xi_{m+1}+\xi_{m+2}+\cdots\xi_{m+n}$两和之间的相关系数。

解．$r_{ST} = E\left\{\left(\sum\limits_{i=1}^{n}(\xi_i - a)\right)\left(\sum\limits_{j=m+1}^{m+n}(\xi_j - a)\right)\right\}\times$

$$\times \frac{1}{\sqrt{n\sigma^2 \cdot n\sigma^2}}$$

$$= \frac{1}{n\sigma^2}\left\{\sum_{i=1}^{m} \sum_{j=m+1}^{n+m} E(\xi_i - a)(\xi_j - a) + \right.$$

$$\left. + \sum_{i=m+1}^{n} E(\xi_i - a)^2 + \sum_{i=m+1}^{n} \sum_{\substack{j=m+1 \\ j\neq i}}^{m+n} E(\xi_i - a)(\xi_j - a)\right\}$$

$$= \frac{(n-m)\sigma^2}{n\sigma^2} = 1 - \frac{m}{n}.$$

24．若ξ的密度函数是偶函数，且$E\xi^2<\infty$，试证$|\xi|$与ξ不相关，但它们不相互独立。

证：设$f(x)$是ξ的密度函数，则$f(-x)=f(x)$。由$xf(x)$是奇函数可得$E\xi=0$，从而$E\xi E|\xi|=0$。又由于

$x|x|f(x)$是奇函数，得

$$E\xi|\xi| = \int_{-\infty}^{\infty} x|x|f(x)dx = 0 = E\xi E|\xi|,$$

故$|\xi|$与ξ不相关。

由于ξ的密度函数是偶函数，故可选$c>0$使$0<P\{|\xi|<c\}<1$，亦有$P\{\xi<c\}<1$，

$$\therefore \quad P\{\xi<c\}P\{|\xi|<c\} \neq P\{|\xi|<c\} = $$
$$= P\{\xi<c, |\xi|<c\}$$

其中等式成立是由于$\{|\xi|<c\}\subset\{\xi<c\}$。由此得$|\xi|$与$\xi$不独立。

25. 若ξ，η的密度函数为

$$p(x,y) = \begin{cases} 1/\pi, & x^2+y^2\leqslant 1 \\ 0, & x^2+y^2>1, \end{cases}$$

试证：ξ与η不相关，但它们不独立。

证：$$E\xi = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xp(x,y)dxdy$$

$$= \int_{-1}^{1} xdx\int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} \frac{1}{\pi}dy = 0,$$

同理 $E\eta = 0$。

$$\mathrm{cov}(\xi,\eta) = E\xi\eta - E\xi E\eta$$

$$= \int_{-1}^{1} xdx\int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} \frac{1}{\pi}ydy = 0,$$

即ξ与η不相关。但ξ与η不独立，事实上可求得

$$p_\xi(x) = \begin{cases} \frac{2}{\pi}\sqrt{1-x^2}, & |x|\leqslant 1 \\ 0, & |x|>1, \end{cases}$$

$$p_\eta(y)=\begin{cases}\frac{2}{\pi}\sqrt{1-y^2}, & |y|\leqslant 1\\ 0, & y>1,\end{cases}$$

而当$|x|\leqslant 1$且$|y|\leqslant 1$时，$p(x,y)\neq p_\xi(x)p_\eta(y)$。

26.若 ξ 与 η 都是只能取两个值的随机变量，试证如果它们不相关，则独立。

证：设$\xi:\begin{pmatrix}a, & b\\ p_1, & q_1\end{pmatrix}$，$\eta:\begin{pmatrix}c, & d\\ p_2, & q_2\end{pmatrix}$。

作两个随机变量

$$\xi^*=\xi-b:\begin{pmatrix}a-b & 0\\ p_1, & q_1\end{pmatrix},\eta^*=\eta-d:\begin{pmatrix}c-d, & 0\\ p_2, & q_2\end{pmatrix}.$$

由 ξ 与 η 不相关即 $E\xi\eta=E\xi E\eta$ 得

$$E\xi^*\eta^*=E(\xi\eta-b\eta-d\xi+bd)=(E\xi E\eta-bE\eta-dE\xi+bd)$$
$$=(E\xi-b)(E\eta-d)=E\xi^*E\eta^*,$$

而 $E\xi^*\eta^*=(a-b)(c-d)P\{\xi^*=a-b,\ \eta^*=c-d\}$，

$E\xi^*E\eta^*=(a-b)P\{\xi^*=a-b\}\cdot(c-d)P\{\eta^*=c-d\}$，

由上两式值相等，再由$(a-b)(c-d)\neq 0$得

$$P\{\xi^*=a-b,\ \eta^*=c-d\}=P\{\xi^*=a-b\}P\{\eta^*=c-d\},$$

此即 $P\{\xi=a,\ \eta=c\}=P\{\xi=a\}\cdot p\{\eta=c\}$。

同理可证

$$P\{\xi=a,\ \eta=d\}=P\{\xi=a\}\cdot P\{\eta=d\},$$
$$P\{\xi=b,\ \eta=c\}=P\{\xi=b\}\cdot P\{\eta=c\},$$
$$P\{\xi=b,\ \eta=d\}=P\{\xi=b\}\cdot P\{\eta=d\},$$

从而 ξ 与 η 独立。

27.若$U=aX+b$，$V=cY+d$，试证 U，V的相关系数等于X，Y的相关系数。

证：$EU=aEX+b$，$DU=a^2DX$，

$$EV = cEY + d,\ DV = c^2 DY,\ \operatorname{cov}(U, V) =$$
$$= Ea(X - EX)c(Y - EY) = ac \cdot \operatorname{cov}(X, Y),$$
$$r_{UV} = \frac{\operatorname{cov}\ (U, V)}{|a|\sqrt{DX}\ |c|\sqrt{D(Y)}} = \frac{ac}{|ac|} r_{xy}.$$

欲 $r_{UV} = r_{xy}$，题中需补设 a 与 c 同号。

28．若 ξ_1, ξ_2, ξ_3 是三个随机变量，试讨论(1) ξ_1, ξ_2, ξ_3 两两不相关；(2)$D(\xi_1 + \xi_2 + \xi_3) = D\xi_1 + D\xi_2 + D\xi_3$；(3)$E\xi_1\xi_2\xi_3 = E\xi_1 \cdot E\xi_2 \cdot E\xi_3$之间的关系。

解：(一)证(1)⇒(2)。设(1)成立，即 ξ_1, ξ_2, ξ_3两两不不相关，则

$$\begin{aligned}
&D(\xi_1 + \xi_2 + \xi_3) \\
&= E[(\xi_1 + \xi_2 + \xi_3) - (E\xi_1 + E\xi_2 + E\xi_3)]^2 \\
&= E[(\xi_1 - E\xi_1) + (\xi_2 - E\xi_2) + (\xi_3 - E\xi_3)]^2 \\
&= D\xi_1 + D\xi_2 + D\xi_3 + 2E(\xi_1 - E\xi_1)(\xi_2 - E\xi_2) + \\
&\qquad + 2E(\xi_1 - E\xi_1)(\xi_3 - E\xi_3) + \\
&\qquad + 2E(\xi_2 - E\xi_2)(\xi_3 - E\xi_3) \\
&= D\xi_1 + D\xi_2 + D\xi_3,
\end{aligned}$$

∴ (2)成立。

(二) (1)⇏(3)。设

$$\xi_1 : \begin{pmatrix} -1, 1 \\ \frac{1}{2}, \frac{1}{2} \end{pmatrix}, \xi_2 : \begin{pmatrix} -1, 1 \\ \frac{1}{2}, \frac{1}{2} \end{pmatrix},$$

并设 ξ_1 与 ξ_2 独立，则

$$\xi_1\xi_2 = \xi_3(\text{记}) : \begin{pmatrix} -1, 1 \\ \frac{1}{2}, \frac{1}{2} \end{pmatrix}, \xi_1\xi_2\xi_3 = \xi_3^2 : \begin{pmatrix} 1 \\ 1 \end{pmatrix},$$

由第三章25题知，ξ_1, ξ_2, ξ_3 两两独立，从而两两不相关，满足(1)。而$E\xi_1 = E\xi_2 = 0$，这时 $E\xi_1 \cdot E\xi_2 \cdot E\xi_3 = 0 \neq 1 = E\xi_1\xi_2\xi_2$，(3)不成立。

(三)(2)$\nRightarrow$(1).设$D\xi>0$，$\xi_1=\xi_2=\xi$，$\xi_3=-\frac{1}{2}\xi$，则

$$D(\xi_1+\xi_2+\xi_3)=D(\xi+\xi-\tfrac{1}{2}\xi)=D(1\tfrac{1}{2}\xi)=2\tfrac{1}{4}D\xi.$$

$$D\xi_1+D\xi_2+D\xi_3=D\xi+D\xi+D(-\tfrac{1}{2}\xi)=2\tfrac{1}{4}D\xi,$$

满足(2).但显然ξ_1,ξ_2,ξ_3两两相关，事实上由$E\xi^2-(E\xi)^2=D\xi\neq0$得$\xi$与$\xi$相关，(1)不成立。

(四)(2)$\nRightarrow$(3).事实上，由(1)$\Leftrightarrow$(2)，(1)$\nRightarrow$(3)得必有(2)$\nRightarrow$(3)。

(五)(3)$\nRightarrow$(2).设

$$\xi_1:\begin{pmatrix}-1,0\\ \frac{1}{2},\frac{1}{2}\end{pmatrix},\ \xi_2=\xi_1+1:\begin{pmatrix}0,1\\ \frac{1}{2},\frac{1}{2}\end{pmatrix},\ \xi_3:\begin{pmatrix}-1,1\\ \frac{1}{2},\frac{1}{2}\end{pmatrix}$$

则

$$\xi_1\xi_2:\begin{pmatrix}0\\ 1\end{pmatrix},\ \xi_1\xi_2\xi_3:\begin{pmatrix}0\\ 1\end{pmatrix}.$$

再设ξ_1与ξ_3独立，从而ξ_1的函数ξ_2与ξ_3也独立，我们有$E\xi_1=-\frac{1}{2}$，$E\xi_2=\frac{1}{2}$，$E\xi_3=0$，$E\xi_1\xi_2=0$，$E\xi_1\xi_3=E\xi_1\cdot E\xi_3=0$，$E\xi_2\xi_3=E\xi_2\cdot E\xi_3=0$，$E\xi_1\xi_2\xi_3=0=E\xi_1\cdot E\xi_2\cdot E\xi_3$，满足(3).但

$$\begin{aligned}D(\xi_1+\xi_2+\xi_3)&=D\xi_1+D\xi_2+D\xi_3+2E\xi_1\xi_2+\\&\quad+2E\xi_1\xi_3+2E\xi_2\xi_3-2E\xi_1\cdot E\xi_2-\\&\quad-2E\xi_1\cdot E\xi_3-2E\xi_2\cdot E\xi_3\\&=D\xi_1+D\xi_2+D\xi_3-2E\xi_1\cdot E\xi_2\\&=D\xi_1+D\xi_2+D\xi_3+\tfrac{1}{2}\neq D\xi_1+D\xi_2+D\xi_3.\end{aligned}$$

∴(2)不成立。

(六)(3)$\nRightarrow$(1).事实上，由(1)$\Leftrightarrow$(2)，(3)$\nRightarrow$(2)得必有(3)$\nRightarrow$(1)。

(七)当ξ_1,ξ_2,ξ_3相互独立时，(1)，(2)，(3)同时成立。

29.若ξ,η服从二元正态分布，$E\xi=a$，$D\xi=1$，$E\eta=b$，

$D\eta=1$.证明:ξ与η的相关系数 $r=\cos q\pi$,其中 $q=P\{(\xi-a)(\eta-b)<0\}$.

证:由题设得

$$q=\frac{1}{2\pi\sqrt{1-r^2}}\iint\limits_{(x-a)(y-b)<0}\exp\left\{-\frac{1}{2(1-r^2)}\times\left[(x-a)^2-2r(x-a)(y-b)+(y-b)^2\right]\right\}dxdy$$

(令 $u=x-a, v=y-b$)

$$=\frac{1}{2\pi\sqrt{1-r^2}}\iint\limits_{uv<0}e^{-\frac{1}{2(1-r^2)}(u^2-2ruv+v^2)}dudv$$

$$=\frac{1}{\pi\sqrt{1-r^2}}\times\int_0^{\infty}dv\int_{-\infty}^0 e^{-\frac{1}{2(1-r^2)}(u^2-2ruv+v^2)}du$$

令 $u=\rho\cos\theta, v=\rho\sin\theta, |J|=\rho$,则

$$q=\frac{1}{\pi\sqrt{1-r^2}}\times\int_{\frac{\pi}{2}}^{\pi}d\theta\int_0^{\infty}\rho e^{-\frac{\rho^2}{2(1-r^2)}(1-r\sin2\theta)}d\rho$$

$$=\frac{1}{\pi\sqrt{1-r^2}}\int_{\frac{\pi}{2}}^{\pi}\left[-\frac{1-r^2}{1-r\sin2\theta}\times e^{-\frac{\rho^2}{2(1-r^2)}(1-r\sin2\theta)}\right]_0^{\infty}d\theta$$

$$=\frac{\sqrt{1-r^2}}{\pi}\int_{\frac{\pi}{2}}^{\pi}\frac{d\theta}{1-r\sin 2\theta}\ (令\alpha=2\theta)$$

$$=\frac{\sqrt{1-r^2}}{2\pi}\int_{\pi}^{2\pi}\frac{d\alpha}{1-r\sin\alpha}$$

$$=\frac{\sqrt{1-r^2}}{\pi}\left[\frac{1}{\sqrt{1-r^2}}\operatorname{arctg}\frac{\operatorname{tg}\frac{1}{2}\alpha-r}{\sqrt{1-r^2}}\right]_{\pi}^{2\pi}$$

由 $\lim\limits_{\alpha\downarrow\pi}\operatorname{tg}\frac{\alpha}{2}=-\infty$，而 $\lim\limits_{\alpha\downarrow\pi}\operatorname{arctg}\frac{\operatorname{tg}\frac{1}{2}\alpha-r}{\sqrt{1-r^2}}=-\frac{\pi}{2}$ 得

$q=\frac{1}{\pi}\operatorname{arctg}\frac{-r}{\sqrt{1-r^2}}+\frac{1}{2}$，即 $\operatorname{tg}\left[\left(q-\frac{1}{2}\right)\pi\right]=\frac{-r}{\sqrt{1-r^2}}$，变形得

$$-\operatorname{ctg}q\pi=-\frac{r}{\sqrt{1-r^2}}，或\operatorname{ctg}^2 q\pi=\frac{r^2}{1-r^2},$$

所以

$$r^2=\frac{\operatorname{ctg}^2 q\pi}{1+\operatorname{ctg}^2 q\pi}=\sin^2 q\pi\cdot\operatorname{ctg}^2 q\pi=\cos^2 q\pi.$$

注意到 $0<q<1$，且 r 与 $\operatorname{ctg}q\pi$ 同号，即 r 与 $\cos q\pi$ 同号，故得 $r=\cos q\pi$（其中 $q=P\{(\xi-a)(\eta-b)<0\}$）。

30. 设 (ξ,η) 服从二元正态分布，$E\xi=E\eta=0$，$D\xi=D\eta=1$，$r_{\xi\eta}=r$，试证

$$E\max(\xi,\eta)=\sqrt{(1-r)/\pi}.$$

证：由题设得

$$E\max(\xi,\eta)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{\max(x,y)}{2\pi\sqrt{1-r^2}}\times$$

$$\times e^{-\frac{1}{2(1-r^2)}(x^2-2xy+y^2)}dxdy$$

$$=\frac{1}{2\pi\sqrt{1-r^2}}\left\{\int_{-\infty}^{\infty}xdx\times\right.$$

$$\times\int_{-\infty}^{x} e^{-\frac{1}{2(1-r^2)}(x^2-2rxy+y^2)}dy+$$

$$+\int_{-\infty}^{\infty}y\,dy\int_{-\infty}^{y} e^{-\frac{1}{2(1-r^2)}(x^2-2rxy+y^2)}dx\Big\}$$

$$=\frac{1}{\pi\sqrt{1-r^2}}\times$$

$$\times\int_{-\infty}^{\infty}x\,dx\int_{-\infty}^{x} e^{-\frac{1}{2(1-r^2)}(x^2-2rxy+y^2)}dy$$

$$=\frac{1}{\pi\sqrt{1-r^2}}\int_{-\infty}^{\infty}xe^{-\frac{x^2}{2}}dx\int_{-\infty}^{x} e^{-\frac{(y-rx)^2}{2(1-r^2)}}dy$$

$$=\frac{1}{\pi}\int_{-\infty}^{\infty}xe^{-\frac{x^2}{2}}dx\int_{-\infty}^{\sqrt{\frac{1-r}{1+r}}x} e^{-\frac{t^2}{2}}dt$$

用分部积分法，令 $dv=xe^{-\frac{1}{2}x^2}dx$，余下部分为 u，得

$E\max(\xi,\eta)=$

$$=\frac{1}{\pi}\Big(-e^{-\frac{x^2}{2}}\int_{-\infty}^{\sqrt{\frac{1-r}{1+r}}x}e^{-\frac{t^2}{2}}dt\Big]_{-\infty}^{\infty}+$$

$$+\frac{1}{\pi}\sqrt{\frac{1-r}{1+r}}\int_{-\infty}^{\infty}e^{-\frac{1}{2}\left(1+\frac{1-r}{1+r}\right)x^2}dx$$

$$=\frac{1}{\pi}\sqrt{\frac{1-r}{1+r}}\int_{-\infty}^{\infty}e^{-\frac{x^2}{1+r}}dx$$

$$= \frac{1}{\pi}\sqrt{\frac{1-r}{1+r}} \cdot \sqrt{2\pi} \cdot \sqrt{\frac{1+r}{2}} = \sqrt{(1-r)/\pi}.$$

31．设 ξ 与 η 独立，具有相同分布 $N(a,\sigma^2)$，试求 $p\xi+q\eta$ 与 $u\xi+v\eta$ 的相关系数。

解：记 $S=p\xi+q\eta, T=u\xi+v\eta$，则

$$ES=pa+qa=(p+q)a, ET=(u+v)a,$$

$$DS=(p^2+q^2)\sigma^2, DT=(u^2+v^2)\sigma^2,$$

$$\text{cov}ST=E(p(\xi-a)+q(\eta-a))\times$$

$$\times(u(\xi-a)-v(\eta-a))=pu\sigma^2+qv\sigma^2,$$

$$\therefore \quad r_{ST}=\frac{pu+qv}{\sqrt{p^2+q^2}\sqrt{u^2+v^2}}.$$

32．Pareto 分布的密度函数为

$$p(x)=\begin{cases} rA^r/x^{r+1}, & x\geqslant A \\ 0, & x<A, \end{cases} \qquad \text{这里} r>0, A>0,$$

试指出这分布具有p阶矩，当且仅当$p<r$。

证：$E\xi^p=\int_A^\infty rA^r x^p \frac{1}{x^{r+1}}\,dx=\int_A^\infty rA^r\frac{1}{x^{r-p+1}}dx.$

当且仅当 $r-p+1>1$，即 $p<r$ 时上式积分收敛，$E\xi^p$ 存在。当 $p<r$ 时，

$$E\xi^p=\frac{r}{r-p}A^r\left(-\frac{1}{x^{r-p}}\right)\Big|_A^\infty=\frac{r}{r-p}A^p.$$

33．若 ξ 的密度函数为

$$p(x)=\begin{cases} \dfrac{1}{2|x|(\log|x|)^2}, & |x|>e \\ 0\quad, & \text{其它}, \end{cases}$$

试证对于任何$\alpha>0$，$E|\xi|^\alpha=\infty$。

证：对$\alpha>0$，由于$\lim\limits_{x\to\infty} x^{\alpha}/(\log x)^2=\infty$，

所以存在$M>e$，使当$x>M$时，$x^{\alpha}/(\log x)^2>1$，此时

$$E|\xi|^{\alpha}=2\int_e^{\infty}\frac{x^{\alpha}}{2x(\log x)^2}dx$$

$$=\int_e^{\infty}\frac{1}{x}\cdot\frac{x^{\alpha}}{(\log x)^2}dx\geqslant\int_M^{\infty}\frac{1}{x}dx=\infty,$$

$\therefore\quad E|\xi|^{\alpha}=\infty$。

34.若ξ服从$N(a,\ \sigma^2)$，试求$E|\xi-a|^k$。

解：$E|\xi-a|^k=\dfrac{1}{\sqrt{2\pi}\sigma}\displaystyle\int_{-\infty}^{\infty}|x-a|^k\times$

$$\times e^{-\frac{(x-a)^2}{2\sigma^2}}dx\left(令\ u=\frac{x-a}{\sigma}\right)$$

$$=\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty}\sigma^k|u|^k e^{-\frac{u^2}{2}}\sigma du=$$

$$=\sqrt{\frac{2}{\pi}}\sigma^k\int_0^{\infty}u^k e^{-\frac{u^2}{2}}du$$

$$=\sqrt{\frac{2}{\pi}}\sigma^k u^{k-1}\left(-e^{-\frac{u^2}{2}}\right)\Big|_0^{\infty}+$$

$$+\sqrt{\frac{2}{\pi}}\sigma^k(k-1)\int_0^{\infty}u^{k-2}e^{-\frac{u^2}{2}}du.$$

当k为偶数时

$E|\xi-a|^k=$

$$=\sqrt{\frac{2}{\pi}}\sigma^k(k-1)(k-3)\cdots3\cdot1\cdot\int_0^{\infty}e^{-\frac{u^2}{2}}du$$

$$=1\cdot3\cdot5\cdots(k-3)(k-1)\sigma^k$$

当 k 为奇数时

$$E|\xi-a|^k=$$

$$=\sqrt{\frac{2}{\pi}}\sigma^k(k-1)(k-3)\cdots4\cdot2\cdot\int_0^{\infty}ue^{-\frac{u^2}{2}}du$$

$$=2\cdot4\cdot\cdots(k-3)(k-1)\sigma^k\sqrt{\frac{2}{\pi}}$$

35. 记$a_k=E|\xi|^k$,若$a_n<\infty$,试证$\sqrt[k]{a_k}\leqslant\sqrt[k+1]{a_{k+1}}$, $k=1,2,\cdots,n-1$.

证:$a_{k-1}t^2+2a_kt+a_{k+1}=$

$$=\int_{-\infty}^{\infty}(t^2|x|^{k-1}+2t|x|^k+|x|^{k+1})dF(x)$$

$$=\int_{-\infty}^{\infty}\left(t|x|^{\frac{1}{2}(k-1)}+|x|^{\frac{1}{2}(k+1)}\right)^2dF(x)\geqslant0$$

即$u(t)=a_{k-1}t^2+2a_kt+a_{k+1}\geqslant0$ 对任意 t 成立。又 $a_{k-1}>0$, 所以判别式 $\triangle=4a_k^2-4a_{k-1}a_{k+1}\leqslant0$,即 $a_k^2\leqslant a_{k-1}a_{k+1}$,

从而有 $a_k^{2k}\leqslant a_{k-1}^k a_{k+1}^k$.

依次令$k=1,2\cdots,n-1$得

$$a_1^2\leqslant a_0a_2,\ a_2^4\leqslant a_1^2a_3^2,\cdots,\ a_{n-1}^{2(n-1)}\leqslant a_{n-2}^{n-1}a_n^{n-1},$$

其中 $a_0=1$.把这些不等式中前 k 个的左右两边分别相乘化简得 $a_k^{k+1}\leqslant a_{k+1}^k$, 两边同开k(k+1)次方,即得

$$\sqrt[k]{a_k} \leqslant \sqrt[k+1]{a_{k+1}}.$$

36.甲袋中装有5只白球，7只黑球，3只红球，乙袋中装有4只白球，4只黑球，7只红球，试问从哪一个袋中取出一只球有较大不肯定性？

解：$H_{甲} = -\frac{5}{15}\log\frac{5}{15} - \frac{7}{15}\log\frac{7}{15} - \frac{3}{15}\log\frac{3}{15}$

$$= \log 15 - \frac{1}{15}(5\log 5 + 7\log 7 + 3\log 3),$$

$H_{乙} = -\frac{4}{15}\log\frac{4}{15} - \frac{4}{15}\log\frac{4}{15} - \frac{7}{15}\log\frac{7}{15}$

$$= \log 15 - \frac{1}{15}(4\log 4 + 4\log 4 + 7\log 7),$$

$\because\quad 5\log 5 + 3\log 3 > 4\log 4 + 4\log 4$

$\therefore\quad H_{乙} > H_{甲}$，乙比甲有较大的不肯定性。

37.试求几何分布的熵。

解：
$$H_\xi = -\sum_{k=1}^{\infty} q^{k-1}p\log q^{k-1}p$$

$$= -\sum_{k=1}^{\infty} q^{k-1}p\log p - p\sum_{k=1}^{\infty}(k-1)q^{k-1}\log q$$

$$= -\log p - p\cdot\frac{q}{(1-q)^2}\log q$$

$$= -\log p - \frac{q}{p}\log q.$$

其中

$$\sum_{k=1}^{\infty}(k-1)q^{k-2} = \left(\sum_{k=1}^{\infty}q^{k-1}\right)' = \left(\frac{1}{1-q}\right)' = \frac{1}{(1-q)^2}$$

38。试求二项分布的熵。

解： $H_\xi = -\sum\limits_{k=0}^{n} C_n^k p^k q^{n-k} \log C_n^k p^k q^{n-k}$

$$= -\sum_{k=0}^{n} C_n^k p^k q^{n-k} \log C_n^k - \sum_{k=0}^{n} k C_n^k p^k q^{n-k} \log p$$

$$- \sum_{k=0}^{n} (n-k) C_n^{n-k} p^k q^{n-k} \log q$$

$$= -\sum_{k=0}^{n} C_n^k p^k q^{n-k} \log C_n^k - np\log p - nq\log q.$$

计算后两项时用到，二项分布$b(k;n,p)$和$b(k;n,q)$ 的数学期望分别为np和nq。

39。若以 α 及 β 分别记二进制信道的输入及输出，已知 $P\{\alpha=1\}=p$，$P\{\alpha=0\}=1-p$，$P\{\beta=1|\alpha=1\}=q$，$P\{\beta=0|\alpha=1\}=1-q$，$P\{\beta=1|\alpha=0\}=r$，$P\{\beta=0|\alpha=0\}=1-r$，试求输出中含有输入的信息量。

解： $P\{\beta=1\}=P\{\beta=1,\ \alpha=0\}+P\{\beta=1,\ \alpha=1\}$

$=P\{\beta=1|\alpha=0\}P\{\alpha=0\}+P\{\beta=1|\alpha=1\}P\{\alpha=1\}$

$=(1-p)r+pq$，

$P\{\beta=0\}=P\{\beta=0,\ \alpha=0\}+P\{\beta=0,\ \alpha=1\}$

$=P\{\beta=0|\alpha=0\}P\{\alpha=0\}+$

$+P\{\beta=0|\alpha=1\}P\{\alpha=1\}$

$=(1-p)(1-r)+(1-q)p$，

$P\{\beta=1\}+P\{\beta=0\}=1$。

$H(出)=H(\beta)=-[(1-p)r+pq]\log[(1-p)r+pq]-$

$-[p(1-q)+(1-p)(1-r)]\times$

$$\times\log[p(1-q)+(1-p)(1-r)],$$

$$H_{\lambda}(出)=H_{\alpha}(\beta)=$$

$$=-P\{\alpha=0\}[P\{\beta=1|\alpha=0\}\log P\{\beta=1|\alpha=0\}+$$

$$+P\{\beta=0|\alpha=0\}\log P\{\beta=0|\alpha=0\}]-$$

$$-P\{\alpha=1\}[P\{\beta=0|\alpha=1\}\log P\{\beta=0|\alpha=1\}+$$

$$+P\{\beta=1|\alpha=1\}\log P\{\beta=1|\alpha=1\}]$$

$$=-(1-p)[r\log r+(1-r)\log(1-r)]-$$

$$-p[(1-q)\log(1-q)+q\log q],$$

所以输出中含有输入的信息量$H(入)-H_{出}(入)$为

$$H(入)-H_{出}(入)=H(出)-H_{\lambda}(出)$$

$$=-[(1-p)r+pq]\log[(1-p)r+pq]-[p(1-q)+$$

$$+(1-p)(1-r)]\log[p(1-q)+(1-p)(1-r)]+$$

$$+(1-p)r\log r+(1-p)(1-r)\log(1-r)+$$

$$+p(1-q)\log(1-q)+pq\log q.$$

40．在12只金属球中混有一只假球，并且不知道它比真球轻还是重，用没有砝码的天平来称这些球，试问至少需要称多少次才能查出这个假球，并确定它比真球轻或重。

解： 需要确定其结局的实验β有24个可能结局，即12个球中有一个是假球，且它比真的轻或重。若认为全部结局是等概的，则实验β的熵$H(\beta)=\log 24$，即需要得到$\log 24$个单位信息。由称一次(随便怎样的)所构成的实验α，可以有3个结局(即天平可以向右斜或向左斜或保持平衡)，进行k次复合试验$A_k=\alpha_1\alpha_2\cdots\alpha_k$后，可得到不大于$k\log 3=\log 3^k$的信息，而$3^2<24<3^3$，所以至少得称三次才可以称出假球，且判明它比真球轻或重。

具体称法共有十几种，详见雅格洛姆著："概率与信

息"，这里仅取一法叙述如下：

第一次称：天平两端分放1234和5678，下余Ⅰ Ⅱ Ⅲ Ⅳ.

(A)若第一次称时平衡，则假球在Ⅰ Ⅱ Ⅲ Ⅳ中。

第二次称：天平两端分放Ⅰ Ⅱ和Ⅲ1， 注意1是真球。

(AA)若第二次称时平衡，则Ⅳ是假球；再把1和Ⅳ分放天平两端称第三次，可判别假球Ⅳ比真球1轻或重。

(AB)若第二次称时Ⅰ Ⅱ较重(或轻)，

第三次称：天平两端分放Ⅰ和Ⅱ。

(ABA)若第三次称时平衡，则Ⅲ是假球，且比真球较轻(或重)。

(ABB)若第三次称时不平衡，则与(AB)中同重(或轻)的那球是假球，且它比真球较重(或轻)。

(B)若第一次称时1234较重，则假球在天平上。

第二次称：天平两端分放125和346。

(BA)若第二次称时平衡，则7,8中之一为假球，由第一次称的结果知假球较轻，再把7和8分放天平两端称第三次，即可找到假球。

(BB)若第二次称时125较重，则或1，2中之一为假球，且它比真球较重，或6是假球且它比真球较轻。

第三次称：天平两端分放1和2。

(BBA)若第三次称时平衡，则6是假球且比真球轻。

(BBB)若第三次称时不平衡，则较重的一球是假球，且它比真球重。

(C)若第一次称时5678较重，则只需把(B)中编号1234与5678依次互换，即得称法。

41. 试用母函数法求巴斯卡分布的数学期望及方差。

解：巴斯卡分布为

$$P\{\xi=n\}=C_{n-1}^{k-1}q^{n-k}p^{k},n=k,k+1,\cdots。$$

其母函数为

$$P(s)=\sum_{n=k}^{\infty}C_{n-1}^{k-1}q^{n-k}p^{k}s^{n}(令m=n-k)$$

$$=p^{k}\sum_{m=0}^{\infty}C_{m+k-1}^{k-1}\ q^{m}s^{m}s^{k}$$

$$=p^{k}s^{k}\sum_{m=0}^{\infty}C_{m+k-1}^{m}(qs)^{m}=\frac{p^{k}s^{k}}{(1-qs)^{k}}。$$

$$E\xi=P'(1)=p^{k}\left(\frac{ks^{k-1}(1-qs)^{k}+ks^{k}(1-qs)^{k-1}q}{(1-qs)^{2k}}\right)_{s=1}$$

$$=\left(\frac{kp^{k}s^{k-1}}{(1-qs)^{k+1}}\right)_{s=1}=\frac{k}{p}。$$

$$P''(1)=\left(\frac{kp^{k}s^{k-1}}{(1-qs)^{k+1}}\right)'_{s=1}$$

$$=kp^{k}\left(\frac{(k-1)s^{k-2}(1-qs)^{k+1}+(k+1)(1-qs)^{k}qs^{k-1}}{(1-qs)^{2k+2}}\right)_{s=1}$$

$$=kp^{k}\cdot\frac{(k-1)(1-q)^{k+1}+(k+1)(1-q)^{k}q}{(1-q)^{2k+2}}$$

$$=\frac{k^{2}+kq}{p^{2}}-\frac{k}{p}。$$

$$D\xi=P''(1)+P'(1)-[P'(1)]^{2}$$

$$=\frac{k^{2}+kq}{p^{2}}-\frac{k}{p}+\frac{k}{p}-\left(\frac{k}{p}\right)^{2}=\frac{kq}{p^{2}}。$$

42。设 ξ 是一个母函数为 $P(s)$ 的随机变量，试求下列各概率对应的母函数 (1)$P\{\xi>n\}$；(2)$P\{\xi=2n\}$。

解：此题指对非负整数 n 的全体而言。

(1)记 $P\{\xi=n\}=p_{n}$，$P\{\xi>n\}=q_{n}$，则

$$q_n = p_{n+1} + p_{n+2} + \cdots,\quad n \geqslant 0。$$

设$P\{\xi > n\}(n \geqslant 0)$的母函数为$Q(s) = \sum_{n=0}^{\infty} q_n s^n$，则当$-1 < s < 1$时，$(1-s)\cdot Q(s)$中$s^n$的系数，当$n \geqslant 1$时为$q_n - q_{n-1} = -p_n$，当$n=0$时为$q_0 = p_1 + p_2 + \cdots = 1 - p_0$，所以

$$(1-s)Q(s) = 1 - p_0 + \sum_{n=1}^{\infty}(-p_n)s^n = 1 - P(s)。$$

所以$Q(s) = [1-P(s)]/(1-s)$，$(-1 < s < 1)$。

(2)$P\{\xi = 2n\}(n \geqslant 0)$的母函数为

$$F(s) = \sum_{n=0}^{\infty} p_{2n} s^n = \frac{1}{2}\left[\sum_{n=0}^{\infty} p_n (s^{\frac{1}{2}})^n + \sum_{n=0}^{\infty} p_n (-s^{\frac{1}{2}})^n\right]$$

$$= \frac{1}{2}\left[P(s^{\frac{1}{2}}) + P(-s^{\frac{1}{2}})\right]。$$

43. 在贝努里试验中，若试验次数v是随机变量，试证成功的次数与失败的次数这两个变量独立的充要条件，是v服从普阿松分布。

证： 设 $\xi_i = \begin{cases} 1, 第\ i\ 次成功 \\ 0, 第\ i\ 次失败, \end{cases}$ $\quad \eta_i = \begin{cases} 1, 第\ i\ 次失败 \\ 0, 第\ i\ 次成功, \end{cases}$

$$P\{\xi_i = 1\} = P\{\eta_i = 0\} = p,\ P\{\xi_i = 0\} = P\{\eta_i = 1\} = 1 - p。$$

则ξ_i的母函数为

$$F_1(s) = (1-p)s^0 + ps^1 = 1 - p + ps。$$

同理可得η_i的母函数为$F_2(s) = p + (1-p)s$。v的母函数记为$G(s)$。以ξ表示成功次数，则$\xi = \xi_1 + \cdots + \xi_v$，本题认为$\{\xi_n\}$与$v$独立，得$\xi$的母函数为$P_\xi(s) = G[F_1(s)]$。同理，以$\eta$表示失败次数，则$\eta = \eta_1 + \cdots + \eta_v$，其母函数为$P_\eta(s) =$
$= G[F_2(s)]$。

必要性。设ξ与η独立，则由$v = \xi + \eta$得

$$G(1-p+ps)\cdot G[p+(1-p)s]=G(s).$$

因为$(1-p+ps)+(p+s-ps)=1+s$，所以若记上式左边G的变量分别为x,y,可得

$$G(x)G(y)=G(x+y-1).$$

令$G(x)=T(x-1)$，则上式变成

$$T(x-1)T(y-1)=T(x+y-1-1)$$
$$=T[(x-1)+(y-1)].$$

利用教本P97引理可得

$$G(x)=T(x-1)=a^{x-1}=e^{\lambda(x-1)},$$

即 v 的母函数$G(s)=\exp\{\lambda(s-1)\}$.这是普阿松分布的母函数。由于母函数与分布列之间是相互唯一确定的，所以得 v 是服从普阿松分布的随机变量。

充分性.设 v 服从普阿松分布，参数为λ，则

$$P\{\xi=r\}=\sum_{n=r}^{\infty}P\{v=n\}P\{\xi_1+\cdots+\xi_\nu=r|v=n\}$$

$$=\sum_{n=r}^{\infty}P\{v=n\}P\{\xi_1+\cdots+\xi_n=r\}$$

$$=\sum_{n=r}^{\infty}\frac{e^{-\lambda}\lambda^n}{n!}\cdot C_n^r p^r(1-p)^{n-r}$$

$$=\frac{e^{-\lambda}}{r!}\sum_{n=r}^{\infty}\frac{1}{(n-r)!}[\lambda(1-p)]^{n-r}(\lambda p)^r$$

$$=e^{-\lambda p}(\lambda p)^r/r!.$$

同理可得

$$P\{\eta=t\}=e^{-\lambda(1-p)}[\lambda(1-p)]^t/t!.$$

又有

$$P\{\xi=r,\eta=t\}=P\{v=r+t\}P\{\xi=r|v=r+t\}.$$

$$= \frac{e^{-\lambda}\lambda^{r+t}}{(r+t)!} \cdot C_{r+t}^{r} p^r (1-p)$$

$$= e^{-\lambda}\lambda^{r+t} p^r (1-p)^t / r!t! = P\{\xi = r\} P\{\eta = t\}.$$

再由r, t的任意性即得证ξ与η独立。

44. 设$\{\xi_k\}$是一串独立的整值随机变量序列，具有相同概率分布，考虑和$\eta = \xi_1 + \xi_2 + \cdots \xi_\nu$，其中$\nu$是随机变量，它与$\{\xi_k\}$相互独立，试用(1)母函数法，(2)直接计算证明

$$E\eta = E\nu \cdot E\xi_k, D\eta = E\nu \cdot D\xi_k + D\nu \cdot (E\xi_k)^2.$$

证： (1)设ξ_i的母函数为$F(s)$，ν的母函数为$G(s)$。而$\eta = \xi_1 + \cdots + \xi_\nu$，所以$P_\eta(s) = G[F(s)]$。由此得

$$E\eta = P'_\eta(1) = \{G'[F(s)] \cdot F'(s)\}_{s=1} = G'(1) \cdot F'(1)$$
$$= E\nu \cdot E\xi_k.$$

其中

$$F(1) = \sum_{j=0}^{\infty} P\{\xi_i = j\} \cdot 1^j = \sum_{j=0}^{\infty} P\{\xi_i = j\} = 1.$$

$$D\eta = P''_\eta(1) + P'_\eta(1) - [P'_\eta(1)]^2$$
$$= \{G''[F(s)] \cdot [F'(s)]^2 + G'[F(s)] \cdot F''(s)\}_{s=1}$$
$$+ F'(1)G'(1) - [F'(1) \cdot G'(1)]^2$$
$$= G''(1)[F'(1)]^2 + G'(1) \cdot F''(1) + F'(1) \cdot G'(1)$$
$$- [F'(1) \cdot G'(1)]^2$$
$$= G'(1)\{F''(1) + F'(1) - [F'(1)]^2\}$$
$$+ [F'(1)]^2 \{G''(1) + G'(1) - [G'(1)]^2\}$$
$$= E\nu \cdot D\xi_k + D\nu \cdot (E\xi_k)^2.$$

(2)直接计算。由题设得

$$P\{\eta = i\} = \sum_{n=0}^{\infty} P\{\nu = n\} P\{\xi_1 + \cdots + \xi_n = i | \nu = n\}.$$

$$E\eta = \sum_{i=0}^{\infty} i \sum_{n=0}^{\infty} P\{\nu = n\} P\{\xi_1 + \cdots + \xi_n = i | \nu = n\}$$

$$=\sum_{n=0}^{\infty}P\{\nu=n\}\sum_{i=0}^{\infty}i\,P\{\xi_1+\cdots+\xi_n=i\}$$

利用$E(\xi_1+\cdots+\xi_n)=nE\xi_1$得

$$E\eta=\sum_{n=0}^{\infty}P\{\nu=n\}nE\xi_1=E\nu\cdot E\xi_1.$$

$$E\eta^2=\sum_{i=0}^{\infty}i^2\sum_{n=0}^{\infty}P\{\nu=n\}P\{\xi_1+\cdots+\xi_n=i|\nu=n\}$$

$$=\sum_{n=0}^{\infty}P\{\nu=n\}\sum_{i=0}^{\infty}i^2P\{\xi_1+\cdots+\xi_n=i\}$$

记$\zeta=\xi_1+\cdots+\xi_n$，利用$E\zeta^2=D\zeta+(E\zeta)^2$及$D\zeta=nD\xi_1$得

$$E\eta^2=\sum_{n=0}^{\infty}P\{\nu=n\}\left(D(\xi_1+\cdots+\xi_n)+[E(\xi_1+\cdots+\xi_n)]^2\right)$$

$$=\sum_{n=0}^{\infty}P\{\nu=n\}[nD\xi_1+n^2(E\xi_1)^2]$$

$$=E\nu\cdot D\xi_1+(E\xi_1)^2\left(\sum_{n=0}^{\infty}n^2P\{\nu=n\}\right)$$

最后，再利用$E\nu^2=D\nu+(E\nu)^2$得

$$E\eta^2=E\nu\cdot D\xi_1+D\nu\cdot(E\xi_1)^2+(E\nu)^2\cdot(E\xi_1)^2.$$

$$\therefore D\eta=E\eta^2-(E\eta)^2=E\nu\cdot D\xi_1+D\nu\cdot(E\xi_1)^2.$$

45. 某公共汽车站在$[0,t]$中来到的乘客批数μ服从参数为λt的普阿松分布，而每批来到的乘客数是随机变量，来n个的概率为p_n，$n=0,1,2,\cdots$，试求$[0,t]$中来到乘客数η的母函数及数学期望。

解： 以ξ_i表第i批来到的乘客数，则ξ_i同分布，且其母函数为$F(s)=\sum_{n=0}^{\infty}p_ns^n$。本题认为，$\xi_i$间相互独立且$\{\xi_n\}$与$\mu$独

立。因为$[0,t]$中来到的乘客数$\eta=\xi_1+\cdots+\xi_\mu$，而$\mu$服从参数为$\lambda t$的普阿松分布，其母函数为$G(s)=\exp\{\lambda t(s-1)\}$，所以$\eta$的母函数为

$$P(s)=G[F(s)]=\exp\{\lambda t(\sum_{n=0}^{\infty}p_n s^n-1)\}.$$

由此得η的数学期望为

$$E\eta=P'(1)=\lambda t\sum_{n=1}^{\infty}np_n.$$

46. 试用母函数法证明二项分布及普阿松分布的再生性。

证：(1)设$\xi_1\sim b(k;n,p)$，$\xi_2\sim b(k;m,p)$，则它们的母函数分别为$P_1(s)=(ps+q)^n$，$P_2(s)=(ps+q)^m$。再设ξ_1与ξ_2独立，$\eta=\xi_1+\xi_2$，则η的母函数为

$$P_\eta(s)=P_1(s)\cdot P_2(s)=(ps+q)^n(ps+q)^m=(ps+q)^{n+m}$$

二项分布$b(k;n+m,p)$的母函数为$(ps+q)^{n+m}$，由于母函数与分布列之间是相互唯一确定的，由此即得η服从$b(k;n+m,p)$，即二项分布具有再生性。

(2)设ξ_1，ξ_2分别服从参数为λ_1，λ_2的普阿松分布，其母函数分别为$P_1(s)=\exp\{\lambda_1(s-1)\}$，$P_2(s)=\exp\{\lambda_2(s-1)\}$。再设$\xi_1$与$\xi_2$独立，$\eta=\xi_1+\xi_2$，则$\eta$的母函数为

$$P_\eta(s)=P_1(s)P_2(s)=\exp\{(\lambda_1+\lambda_2)(s-1)\}.$$

所以η服从参数为$\lambda_1+\lambda_2$的普阿松分布，普阿松分布具有再生性。

47. 若分布函数$F(x)=1-F(-x+0)$成立，则称它是**对称的**。试证分布函数对称的充要条件，是它的特征函数是实的偶函数。

证：必要性。由$F(x)=1-F(-x+0)$得

$$P\{\xi<x\}=P\{\xi>-x\}=P\{-\xi<x\},$$

此即$F_\xi(x)=F_{-\xi}(x)$，所以对特征函数$f(t)$有

$$f(t)=Ee^{it\xi}=\int e^{itx}dF_\xi(x)=\int e^{itx}dF_{-\xi}(x)=Ee^{-it\xi}$$
$$=\overline{f(t)},$$

由此知$f(t)$是实函数。又有

$$f(-t)=\int e^{-itx}dF_\xi(x)=\int e^{-itx}dF_{-\xi}(x)=Ee^{-it(-\xi)}$$
$$=Ee^{it\xi}=f(t),$$

所以$f(t)$又是偶函数。

充分性。由于

$\overline{f_\xi(t)}=Ee^{-it\xi}=Ee^{it(-\xi)}=f_{-\xi}(t)$，

又由题设知$f_\xi(t)$是实函数，所以$f_\xi(t)=\overline{f_\xi(t)}=f_{-\xi}(t)$。由唯一性定理知，$\xi$与$-\xi$的分布函数相同，$F_\xi(x)=F_{-\xi}(x)$，即

$P\{\xi<x\}=P\{-\xi<x\}=P\{\xi>-x\}$，

从而$F(x)=1-F(-x+0)$。

48。试求[0,1]均匀分布的特征函数。

解： $p_\xi(x)=\begin{cases}1, & x\in[0,1]\\ 0, & x\notin[0,1].\end{cases}$

当$t=0$时$f(t)=1$；当$t\neq0$时

$$f(t)=\int_0^1 e^{itx}dx=\frac{1}{it}e^{itx}\Big|_0^1=\frac{1}{it}(e^{it}-1).$$

49。一般柯西分布的密度函数为

$$p(x)=\frac{1}{\pi}\cdot\frac{\lambda}{\lambda^2+(x-\mu)^2},\quad \lambda>0.$$

试证它的特征函数为$\exp\{i\mu t-\lambda|t|\}$，利用这个结果证明柯西分布的再生性。

证：$f(t)=\int_{-\infty}^{\infty}e^{itx}\cdot\frac{1}{\pi}\cdot\frac{\lambda}{\lambda^2+(x-\mu)^2}dx\left(令u=\frac{x-\mu}{\lambda}\right)$

$$=\frac{1}{\pi}e^{it\mu}\int_{-\infty}^{\infty}e^{it\lambda u}\frac{1}{1+u^2}du. \qquad (1)$$

考虑复变函数的积分，当$t>0$时，取 c 为上半圆周 $\wp=Re^{i\theta}(0\leqslant\theta\leqslant\pi)$和实轴上从$-R$到$R$的围道(图15)，若u 位于上半圆周上，则$u=Re^{i\theta}$，$du=iRe^{i\theta}d\theta$，有

$$\int_C e^{it\lambda z}\frac{1}{1+z^2}dz=\int_{-R}^{R}e^{it\lambda u}\frac{1}{1+u^2}du+$$

$$+\int_0^{\pi}iRe^{i\theta}\frac{e^{it\lambda Re^{i\theta}}}{1+R^2e^{2i\theta}}d\theta=I_1+I_2. \qquad (2)$$

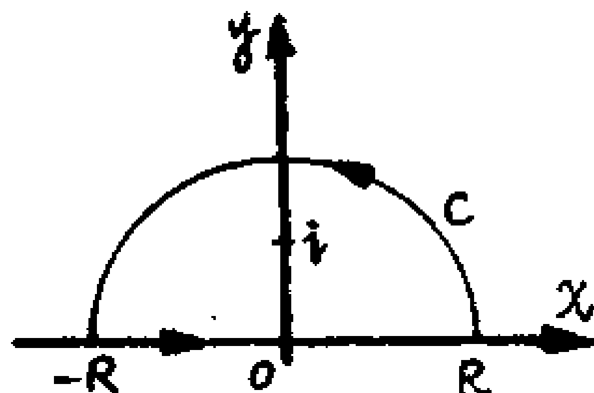


图 15

对I_1有

$$\lim_{R\to\infty}I_1=\int_{-\infty}^{\infty}e^{it\lambda u}\frac{1}{1+u^2}du.$$

由$t>0$及题设$\lambda>0$得$0<e^{-t\lambda R\sin\theta}\leqslant1$，所以对$I_2$有

$$|I_2|\leqslant R\int_0^{\pi}\left|\frac{e^{it\lambda Re^{i\theta}}}{1+R^2e^{2i\theta}}\right|d\theta\leqslant\frac{R}{R^2-1}\int_0^{\pi}e^{-\lambda tR\sin\theta}d\theta$$

$$<\frac{R\pi}{R^2-1}\to0(当R\to+\infty时) \qquad (3)$$

在上半平面上，仅有$z=i$是被积函数的一阶极点，由复变函数中留数定理得，对任何$R>1$有

$$\int_C e^{it\lambda z}\frac{1}{1+z^2}dz=2\pi i\cdot c_{-1}=2\pi i\frac{e^{-\lambda t}}{2i}=\pi e^{-\lambda t}, \quad (4)$$

其中

$$c_{-1}=\lim_{z\to i}\frac{e^{i\lambda tz}}{1+z^2}(z-i)=\lim_{z\to i}\frac{e^{i\lambda tz}}{(z+i)(z-i)}(z-i)$$

$$=\lim_{z\to i}\frac{e^{i\lambda tz}}{z+i}=\frac{1}{2i}e^{i\lambda ti}=\frac{1}{2i}e^{-\lambda t}.$$

把(2),(3),(4)代入(1)式得

$$\int_{-\infty}^{\infty}e^{it\lambda u}\frac{1}{1+u^2}du=\pi e^{-\lambda t}. \quad (5)$$

由于

$$e^{it\lambda u}\frac{1}{1+u^2}=\frac{1}{1+u^2}\cos t\lambda u+\frac{i}{1+u^2}\sin t\lambda u,$$

$\sin t\lambda u/(1+u^2)$是u的奇函数，它在$(-\infty,\infty)$上积分值为0；$\cos t\lambda u/(1+u^2)$是t的偶函数，当$t<0$时，其积分值应与$t>0$时积分值相等；再注意到(5)中右端$t>0$，所以当$t<0$时有

$$\int_{-\infty}^{\infty}e^{it\lambda u}\frac{1}{1+u^2}du=\pi e^{-\lambda|t|} \quad (6)$$

当$t=0$时有

$$\int_{-\infty}^{\infty}e^{it\lambda u}\frac{1}{1+u^2}du=\int_{-\infty}^{\infty}\frac{1}{1+u^2}du=\pi=\pi e^{-\lambda\cdot 0} \quad (7)$$

把(5)—(7)代入(1)式得，对任意t有

$$f(t)=\exp\{it\mu-\lambda|t|\}.$$

现证柯西分布具有再生性。设$\xi_i(i=1,2)$的特征函数为$f_i(t)=\exp\{it\mu_i-\lambda_i|t|\}$，再设$\xi_1$与$\xi_2$独立，$\eta=\xi_1+\xi_2$，则

$$f_\eta(t)=f_1(t)f_2(t)=\exp\{it(\mu_1+\mu_2)-(\lambda_1+\lambda_2)|t|\},$$

所以 η 仍服从柯西分布，且参数为$\mu_1+\mu_2,\lambda_1+\lambda_2$。

50. 若随机变量 ξ 服从柯西分布，$\mu=0,\lambda=1$，而$\eta=\xi$，试证关于特征函数成立着 $f_{\xi+\eta}(t)=f_\xi(t)\cdot f_\eta(t)$，但是 ξ与 η 并不独立。

证：由上题得$f_\xi(t)=f_\eta(t)=e^{-|t|}$，所以由$\xi+\eta=2\xi$ 得

$$f_{\xi+\eta}(t)=f_{2\xi}(t)=e^{-|2t|}=e^{-2|t|}=f_\xi(t)\cdot f_\eta(t).$$

但ξ与η并不独立，事实上，可取 c 使$0<P\{\xi<c\}<1$，则

$$P\{\xi<c,\eta<c\}=P\{\xi<c\}\neq P\{\xi<c\}\cdot P\{\eta<c\},$$

这说明由 ξ 与 η 独立可推得$f_{\xi+\eta}(t)=f_\xi(t)\cdot f_\eta(t)$，但反之不真。

51. 试求指数分布与 Γ- 分布的特征函数，并证明对于具有相同 λ 值的 Γ- 分布，关于参数 r 有再生性。

解：(1)指数分布。当$x\geqslant0$时，其密度函数为

$p(x)=\lambda e^{-\lambda x}$，所以它的特征函数为

$$f(t)=\int_0^\infty e^{itx}\lambda e^{-\lambda x}dx=\int_0^\infty \lambda e^{x(-\lambda+it)}dx$$

$$=\frac{\lambda}{it-\lambda}e^{x(it-\lambda)}\Big|_0^\infty=\frac{\lambda}{\lambda-it}=\left(1-\frac{it}{\lambda}\right)^{-1},$$

其中 $|e^{x(it-\lambda)}|=e^{-\lambda x}\to0(x\to+\infty)$。

(2)Γ- 分布。当 $x>0$时，其密度函数为

$$p(x)=\frac{\lambda^r}{\Gamma(r)}x^{r-1}e^{-\lambda x},$$

为求其特征函数，我们指出，对复数$z=b+ic$，只要$b>0$，就有如下等式成立（参阅克拉美著"统计学数学方法"（魏宗舒译）P124）：

$$\int_0^\infty x^{r-1}e^{-zx}dx=\frac{\Gamma(r)}{z^r}.$$

利用此式可求得Γ-分布的特征函数为

$$f(t)=\int_0^{\infty} e^{itx}\frac{\lambda^r}{\Gamma(r)}x^{r-1}e^{-\lambda x}dx$$

$$=\frac{\lambda^r}{\Gamma(r)}\int_0^{\infty} x^{r-1}e^{-(\lambda-it)x}dx$$

$$=\frac{\lambda^r}{\Gamma(r)}\cdot\frac{\Gamma(r)}{(\lambda-it)^r}=\left(1-\frac{it}{\lambda}\right)^{-r}.$$

现证Γ-分布具有再生性。设$\xi_1\sim G(\lambda,r_1)$，$\xi_2\sim G(\lambda,r_2)$，则它们的特征函数分别为

$$f_1(t)=(1-it/\lambda)^{-r_1},\quad f_2(t)=(1-it/\lambda)^{-r_2},$$

再设ξ_1与ξ_2独立，$\eta=\xi_1+\xi_2$，则有

$$f_\eta(t)=f_1(t)\cdot f_2(t)=(1-it/\lambda)^{-r_1-r_2},$$

所以η服从Γ-分布$G(\lambda,r_1+r_2)$，Γ-分布具有再生性。

52. 若$\xi_1,\xi_2,\cdots,\xi_n$相互独立且服从相同分布$N(0,1)$，试证$\eta=\sum\limits_{i=1}^{n}\xi_i^2$服从参数为$n$的$\chi^2$-分布，并说明$\chi^2$-分布也有再生性。

证： 记ξ_i^2的分布函数为$F(y)$，则当$y\leqslant0$时$F(y)=0$；当$y>0$时

$$F(y)=P\{-\sqrt{y}<\xi_i<\sqrt{y}\}$$

$$=\int_{-\sqrt{y}}^{\sqrt{y}}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}dx.$$

利用对参变量积分求导法则，对$F(y)$求导可得ξ_i^2的分布密度$p(y)$，当$y\leqslant0$时$p(y)=0$；当$y>0$时

$$p(y)=\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}y}\left(\frac{1}{2\sqrt{y}}+\frac{1}{2\sqrt{y}}\right)$$

$$=\frac{(\frac{1}{2})^{\frac{1}{2}}}{\Gamma(\frac{1}{2})}y^{\frac{1}{2}-1}e^{-\frac{1}{2}y}.$$

把此式与 χ^2- 分布密度（见第三章35题解答）比较可知，ξ_i^2 服从自由度为 1 的 χ^2- 分布，也就是服从Γ- 分布 $G(\frac{1}{2},\frac{1}{2})$。由 ξ_i 间独立得 ξ_i^2 间也独立，利用上题结论可得 $\eta=\sum\limits_{i=1}^{n}\xi_i^2$ 服从Γ- 分布 $G(\frac{1}{2},\ \frac{1}{2}n)$，即自由度为n 的 χ^2-分布。再由上题中Γ- 分布具有再生性可得，这里 χ^2- 分布也具有再生性。

53。求证：对于任何实值特征函数f(t)，以下两个不等式成立：

$1-f(2t)\leqslant4(1-f(t))$，$1+f(2t)\geqslant2(f(t))^2$。

证：f(t)是实值函数，复数部分为0，只需对实部计算。

$$1-f(2t)=\int_{-\infty}^{\infty}(1-\cos2tx)dF(x)$$

$$=\int_{-\infty}^{\infty}2\sin^2tx\,dF(x)$$

$$=2\int_{-\infty}^{\infty}(1-\cos tx)(1+\cos tx)dF(x)$$

$$\leqslant4\int_{-\infty}^{\infty}(1-\cos tx)dF(x)=4(1-f(t)).$$

$$1+f(2t)=\int_{-\infty}^{\infty}(1+\cos2tx)dF(x)$$

$$=2\int_{-\infty}^{\infty}\cos^2tx\,dF(x)$$

$$\geqslant 2\left(\int_{-\infty}^{\infty} \cos tx dF(x)\right)^2 = 2(f(t))^2,$$

其中利用柯西—许瓦兹不等式（置$\varphi(x)=\cos tx$，$\psi(x)=1$）

$$\left[\int \varphi(x)\psi(x)dF(x)\right]^2 \leqslant$$

$$\leqslant \left(\int \varphi^2(x)dF(x)\right)\left(\int \psi^2(x)dF(x)\right)$$

54．求证：如果$f(t)$是相应于分布函数$F(x)$的特征函数，则对于任何x值恒成立：

$$\lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T} f(x)e^{-itx}dt = F(x+0)-F(x-0).$$

证：$$I = \lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T} f(t)e^{-itx}dt$$

$$= \lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T}\left(\int_{-\infty}^{\infty} e^{it(y-x)}\, dF(y)\right)dt.$$

由于$|e^{it(y-x)}|=1$，所以$e^{it(y-x)}$关于乘积测度

$P_F\times L[-T,T]$绝对可积，由富比尼定理知可交换上式中积分次序，得

$$I = \lim_{T\to\infty} \frac{1}{2T}\int_{-\infty}^{\infty} dF(y)\left(\int_{-T}^{T} e^{it(y-x)}dt\right).$$

记$g(T,y)=\frac{1}{2T}\int_{-T}^{T} e^{it(y-x)}dt$，则当$y=x$时有

$$g(T,y)=\frac{1}{2T}\int_{-T}^{T} dt = 1,$$

当$y\neq x$时有

$$g(T,y)=\frac{1}{2T}\int_{0}^{T} 2\cos t(y-x)dt = \frac{\sin T(y-x)}{T(y-x)}.$$

由此得$|g(T,y)|\leqslant 1$，且$\lim\limits_{T\to\infty} g(T,y)=\begin{cases}0, & y\neq x\\ 1, & y=x.\end{cases}$

由控制收敛定理得

$$I=\int_{-\infty}^{\infty}\left(\lim_{T\to\infty} g(T,y)\right)dF(y)=\int_{\{x\}} dF(y)$$

$$=P\{x\}=F(x+0)-F(x).$$

55. 随机变量ξ的特征函数为$f(t)$，且它的 n 阶矩存在，令

$$\chi_k=\frac{1}{i^k}\left[\frac{d^k}{dt^k}\log f(t)\right]_{t=0},\quad k\leqslant n,$$

称χ_k为随机变量ξ的k**阶半不变量**.试证$\eta=\xi+b$(b是常数)的k(k>1)阶半不变量等于χ_k.

证： 由$\eta=\xi+b$得$f_\eta(t)=e^{itb}f_\xi(t)$，亦有

$$\log f_\eta(t)=itb+\log f_\xi(t).$$

当 k>1 时，等式两边同对 t 求 k 阶导数，itb一项导数为 0 所以由定义得 η 的 χ_k 等于 ξ 的 χ_k.

56. 试求出半不变量与原点矩之间的关系式.

解： 利用特征函数$f(t)$与原点矩m_k之间的关系式$f^{(k)}(0)=i^k m_k$，可把$f(t)$展成幂级数

$$f(t)=1+\sum_{k=1}^{\infty}\frac{m_k}{k!}(it)^k,\quad f(0)=1. \tag{1}$$

又利用上题中定义的χ_k，可把$\log f(t)$展成幂级数

$$\log f(t)=\sum_{k=1}^{\infty}\frac{\chi_k}{k!}(it)^k,$$

$$\therefore\ f(t)=\exp\left\{\sum_{k=1}^{\infty}\frac{\chi_k}{k!}(it)^k\right\}. \tag{2}$$

再把(2)中的e^y展成幂级数得

$$f(t)=1+\frac{1}{1!}\sum_{k=1}^{\infty}\frac{\chi_k}{k!}(it)^k+$$

$$+\frac{1}{2!}\left[\sum_{k=1}^{\infty}\frac{\chi_k}{k!}(it)^k\right]^2+\cdots。\qquad(3)$$

比较(1)与(3)式中$(it)^k$的系数，可得半不变量与原点矩之间的关系式

$$m_1=\chi_1, m_2=\chi_2+\chi_1^2, m_3=\chi_3+3\chi_1\chi_2+\chi_1^3,\cdots。$$

57．若(ξ,η)服从二元正态分布，试求$\zeta=\xi+\eta$的分布密度函数。

解：利用教本 P222 定理 6：$\xi=(\xi_1,\cdots,\xi_n)^T$服从n元正态分布$N(a,B)$的充要条件，是它的任何一个线性组合

$$\zeta=\sum_{j=1}^{n}l_j\xi_j\text{服从一元正态分布}N\left(\sum_{j=1}^{n}l_ja_j,\ \sum_{j,k=1}^{n}l_jl_kb_{jk}\right),$$

其中$a_k=E\xi_k, b_{jk}=E(\xi_j-E\xi_j)(\xi_k-E\xi_k)$。

设$(\xi,\eta)\sim N(a,B)$ 其中

$$a=\begin{pmatrix}a_1\\a_2\end{pmatrix}, B=\begin{pmatrix}\sigma_1^2 & r\sigma_1\sigma_2\\ r\sigma_1\sigma_2 & \sigma_2^2\end{pmatrix}=\begin{pmatrix}b_{11} & b_{12}\\ b_{21} & b_{22}\end{pmatrix}.$$

由上述定理 6 得，$\zeta=\xi+\eta\sim N(a_1+a_2, \sigma_1^2+2r\sigma_1\sigma_2+\sigma_2^2)$，所以其分布密度函数为

$$p_\zeta(z)=\frac{1}{\sqrt{2\pi}\sqrt{\sigma_1^2+2r\sigma_1\sigma_2+\sigma_2^2}}\times$$

$$\times\exp\left\{-\frac{(z-a_1-a_2)^2}{2(\sigma_1^2+\sigma_2^2+2r\sigma_1\sigma_2)}\right\}.$$

58．设$\xi_1,\xi_2,\cdots,\xi_n$相互独立，具有相同分布$N(a,\sigma^2)$，

试求$\xi=\begin{pmatrix}\xi_1\\ \vdots\\ \xi_n\end{pmatrix}$的分布，并写出它的数学期望及协方差阵，再求$\overline{\xi}=\frac{1}{n}\sum_{i=1}^{n}\xi_i$的分布密度。

解： 由诸ξ_i独立得ξ的密度函数为

$$p(x_1,\cdots,x_n)=\prod_{i=1}^{n}p_{\xi_i}(x)$$

$$=\frac{1}{(\sqrt{2\pi})^n\sigma^n}\prod_{i=1}^{n}\exp\left\{-\frac{(x_i-a)^2}{2\sigma^2}\right\},$$

数学期望和协方差阵为

$$E\xi=\begin{pmatrix}a\\ \vdots\\ a\end{pmatrix}_{n\times1},\qquad B=\begin{pmatrix}\sigma^2 & & & 0\\ & \sigma^2 & & \\ & & \ddots & \\ 0 & & & \sigma^2\end{pmatrix}_{n\times n}.$$

由上题解答中叙述的定理 6 知，

$$\overline{\xi}\sim N\left(\frac{1}{n}\sum_{i=1}^{n}a,\ \frac{1}{n^2}\sum_{i=1}^{n}\sigma^2\right)=N\left(a,\ \frac{\sigma^2}{n}\right),$$

所以$\overline{\xi}$的分布密度为

$$p_{\overline{\xi}}(x)=\frac{\sqrt{n}}{\sqrt{2\pi}\,\sigma}\exp\left\{-\frac{n(x-a)^2}{2\sigma^2}\right\}.$$

59. 若ξ服从二元正态分布$N(0,\Sigma)$，其中$\Sigma=\begin{pmatrix}4&2\\2&1\end{pmatrix}$，试找出矩阵 A，使$\xi=A\eta$，且要求$\eta$服从非退化的正态分布，并求$\eta$的密度函数。

解： 取$E\eta=0$，$D\eta=C$，$\eta\sim N(0,C)$，$C=\begin{pmatrix}1&0\\0&1\end{pmatrix}$.

令$\xi = A\eta$，其中$A=(a_{ij})_{2\times 2}$，则η与ξ的特征函数分别为

$$f_\eta(t)=\exp\{-\tfrac{1}{2}t^\tau Ct\},f_\xi(t)=\exp\{-\tfrac{1}{2}t^\tau(ACA^\tau)t\},$$

且有$ACA^\tau=\Sigma$，即

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \end{pmatrix}$$

$$=\begin{pmatrix} a_{11}^2+a_{12}^2 & a_{11}a_{21}+a_{12}a_{22} \\ a_{11}a_{21}+a_{12}a_{22} & a_{21}^2+a_{22}^2 \end{pmatrix}=\begin{pmatrix} 4 & 2 \\ 2 & 1 \end{pmatrix}.$$

矩阵A不唯一，取$a_{11}=\sqrt{2}$可解得$a_{12}=\sqrt{2}$，$a_{21}=\sqrt{2}/2$，$a_{22}=\sqrt{2}/2$，从而$A=\begin{pmatrix} \sqrt{2} & \sqrt{2} \\ \sqrt{2}/2 & \sqrt{2}/2 \end{pmatrix}$，这时满足题中的要求。由$\eta\sim N(0,C)$得$\eta$非退化，且$\eta$的密度函数为

$$p(x_1,x_2)=\frac{1}{2\pi}\exp\left\{-\frac{1}{2}(x_1^2+x_2^2)\right\}.$$

60．证明：在正交变换下，多元正态分布的独立、同方差性不变。

证：设$\xi=(\xi_1,\cdots,\xi_n)^\tau$，$\xi_i$独立同方差，其协方差矩阵和特征函数分别为

$$D\xi=\begin{pmatrix} \sigma^2 & & 0 \\ & \ddots & \\ 0 & & \sigma^2 \end{pmatrix}=B,$$

$$f_\xi(t)=\exp\{ia^\tau t-\tfrac{1}{2}t^\tau Bt\}.$$

再设$\eta=C\xi$，其中$C=\begin{pmatrix} c_{11}\cdots c_{1n} \\ \cdots\cdots\cdots \\ c_{n1}\cdots c_{nn} \end{pmatrix}$是正交矩阵，即满足

$$\sum_{j=1}^{n} c_{ij}^2 = 1, \sum_{j=1}^{n} c_{ji}c_{ki} = 0 \quad (j \neq k).$$

由此得$\eta \sim N(\mathbf{Ca}, \mathbf{CBC}^\tau)$，其特征函数为

$$f_\eta(t) = \exp\{i(\mathbf{Ca})^\tau t - \frac{1}{2} t^\tau (\mathbf{CBC}^\tau) t\},$$

即 η 的协方差矩阵为$\mathbf{CBC}^\tau$。利用$\mathbf{C}$的正交性计算得

$$\mathbf{CBC}^\tau = \begin{pmatrix} c_{11} \cdots c_{1n} \\ \cdots\cdots\cdots \\ c_{n1} \cdots c_{nn} \end{pmatrix} \begin{pmatrix} \sigma^2 & & 0 \\ & \ddots & \\ 0 & & \sigma^2 \end{pmatrix} \begin{pmatrix} c_{11} \cdots c_{n1} \\ \cdots\cdots\cdots \\ c_{1n} \cdots c_{nn} \end{pmatrix}$$

$$= \begin{pmatrix} \sigma^2 c_{11} \cdots \sigma^2 c_{1n} \\ \cdots\cdots\cdots\cdots \\ \sigma^2 c_{n1} \cdots \sigma^2 c_{nn} \end{pmatrix} \begin{pmatrix} c_{11} \cdots c_{n1} \\ \cdots\cdots\cdots \\ c_{1n} \cdots c_{nn} \end{pmatrix}$$

$$= \begin{pmatrix} \sum c_{1i}^2 \sigma^2 & \sum c_{1i}c_{2i}\sigma^2 & \cdots & \sum c_{1i}c_{ni}\sigma^2 \\ \sum c_{2i}c_{1i}\sigma^2 & \sum c_{2i}^2 \sigma^2 & \cdots & \sum c_{2i}c_{ni}\sigma^2 \\ \cdots\cdots & \cdots\cdots & \cdots\cdots & \cdots\cdots \\ \sum c_{ni}c_{1i}\sigma^2 & \sum c_{ni}c_{2i}\sigma^2 & \cdots & \sum c_{ni}^2 \sigma^2 \end{pmatrix}$$

$$= \begin{pmatrix} \sigma^2 & & 0 \\ & \ddots & \\ 0 & & \sigma^2 \end{pmatrix} = \mathbf{B},$$

矩阵中$\sum$都是对 i 从 1 到 n 求和的。由协方差矩阵知，η 的各分量 $\eta_1, \cdots, \eta_n$ 间两两不相关且同方差，再由正态分布 间相互独立的充要条件是它们两两不相关 得，$\eta_1, \cdots, \eta_n$ 相 互独立且同方差。

第五章　极限定理

本章内容　贝努里试验场合的极限定理，各种收敛性，独立同分布场合的极限定理，强大数定律，中心极限定理。

1. 设$f(x)(0\leqslant x<\infty)$是单调非降函数，且$f(x)>0$，对随机变量$\xi$，若$Ef(|\xi|)<\infty$，则对任意$x>0$，

$$P\{|\xi|\geqslant x\}\leqslant\frac{1}{f(x)}Ef(|\xi|).$$

证： 对任意$x>0$，

$$P\{|\xi|\geqslant x\}=\int_{|y|\geqslant x}dF(y)\leqslant\frac{1}{f(x)}\int_{|y|\geqslant x}f(|y|)dF(y)$$

$$\leqslant\frac{1}{f(x)}\int_{-\infty}^{\infty}f(|y|)dF(y)=\frac{1}{f(x)}Ef(|\xi|)$$

2. ξ为非负随机变量，若$Ee^{a\xi}<\infty(a>0)$，则对任意$x>0$，

$$P\{\xi\geqslant x\}\leqslant e^{-ax}Ee^{a\xi}.$$

证： 对任意$x>0$，

$$P\{\xi \geqslant x\} = \int_{y\geqslant x} dF(y) \leqslant \frac{1}{e^{ax}} \int_{y\geqslant x} e^{ay} dF(y)$$

$$\leqslant \frac{1}{e^{ax}} \int_0^\infty e^{ay} dF(y) = e^{-ax} E e^{a\xi}$$

3. 若$h(x) \geqslant 0$，ξ为随机变量，且$Eh(\xi) < \infty$，则关于任何$c > 0$，$P\{h(\xi) \geqslant c\} \leqslant c^{-1} Eh(\xi)$。

证： $h(\xi)$为非负随机变量，所以对任意$c > 0$有

$$P\{h(\xi) \geqslant c\} = \int_{x\geqslant c} dF_h(x) \leqslant \frac{1}{c} \int_{x\geqslant c} x dF_h(x)$$

$$\leqslant \frac{1}{c} \int_0^\infty x dF_h(x) = \frac{1}{c} Eh(\xi).$$

4. $\{\xi_k\}$各以$\frac{1}{2}$概率取值k^s和$-k^s$，当 s 为何值时，大数定律可用于独立随机变量序列$\xi_1, \cdots, \xi_n, \cdots$的算术平均值？

解： 现验证何时满足马尔可夫条件$\frac{1}{n^2} D(\sum_{k=1}^n \xi_k) \to 0$

$$E\xi_k = \frac{1}{2}k^s - \frac{1}{2}k^s = 0, \quad D\xi_k = \frac{1}{2}k^{2s} + \frac{1}{2}k^{2s} = k^{2s}.$$

若$s < \frac{1}{2}$，这时$2s - 1 < 0$，利用ξ_k间的独立性可得

$$\frac{1}{n^2} D\left(\sum_{k=1}^n \xi_k\right) = \frac{1}{n^2} \sum_{k=1}^n k^{2s} \leqslant \frac{n \cdot n^{2s}}{n^2} =$$

$$= n^{2s-1} \to 0 (n \to \infty).$$

若$s \geqslant \frac{1}{2}$，则

$$\frac{1}{n^2}D\left(\sum_{k=1}^{n}\xi_k\right)=\frac{1}{n^2}\sum_{k=1}^{n}k^{2s}\geqslant\frac{1}{n^2}\sum_{k=1}^{n}k=$$

$$=\frac{n+1}{2n}\nrightarrow 0(n\rightarrow\infty)。$$

所以当$s<\frac{1}{2}$时，大数定律可用于独立随机变量序列$\{\xi_n\}$。

5．若ξ_k的分布列为$\begin{pmatrix}\sqrt{\ln k}, & -\sqrt{\ln k}\\ \frac{1}{2}, & \frac{1}{2}\end{pmatrix}$，试证大数定律适用于独立随机变量序列$\{\xi_k\}$。

证： $E\xi_k=0$，

$$D\xi_k=\frac{1}{2}(\sqrt{\ln k})^2+\frac{1}{2}(\sqrt{\ln k})^2=\ln k,$$

$$\frac{1}{n^2}D\left(\sum_{k=1}^{n}\xi_k\right)=\frac{1}{n^2}\sum_{k=1}^{n}\ln k\leqslant$$

$$\leqslant\frac{n\ln n}{n^2}$$

$$=\frac{\ln n}{n}\longrightarrow 0(n\rightarrow\infty)。$$

满足马尔可夫条件，故题中结论成立。

6．验证概率分布如下给定的独立随机变量序列是否满足

马尔可夫条件：

(1) $P\{X_k=\pm 2^k\}=\frac{1}{2}$；

(2) $P\{X_k=\pm 2^k\}=2^{-(2k+1)}$, $P\{X_k=0\}=1-2^{-2k}$；

(3) $P\{X_k=\pm k\}=\frac{1}{2}k^{-\frac{1}{2}}$, $P\{X_k=0\}=1-k^{-\frac{1}{2}}$.

证：(1) $EX_k=0$, $DX_k=(2^k)^2\cdot\frac{1}{2}+(-2^k)^2\cdot\frac{1}{2}=4^k$,

$$\frac{1}{n^2}D\left(\sum_{k=1}^{n}X_k\right)=\frac{1}{n^2}\sum_{k=1}^{n}DX_k=\frac{1}{n^2}\times\frac{4-4^{n+1}}{1-4}$$

$$=\frac{4^{n+1}}{3n^2}-\frac{4}{3n^2}\to\infty(n\to\infty).$$

不满足马尔可夫条件。

(2) $EX_k=0$,

$$DX_k=(2^k)^2\cdot\frac{1}{2^{2k+1}}+(-2^k)^2\cdot\frac{1}{2^{2k+1}}=1,$$

$$\frac{1}{n^2}D\left(\sum_{k=1}^{n}X_k\right)=\frac{1}{n^2}\cdot n=\frac{1}{n}\to 0(n\to\infty).$$

满足马尔可夫条件。

(3) $EX_k=0$, $DX_k=k^2\cdot\frac{1}{2k^{1/2}}+(-k)^2\cdot\frac{1}{2k^{1/2}}$

$$=k^{3/2},$$

$$\frac{1}{n^2}D\left(\sum_{k=1}^{n}X_k\right)=\frac{1}{n^2}\sum_{k=1}^{n}k^{3/2}\geqslant\frac{1}{n^2}\sum_{k=1}^{n}k$$

$$=\frac{1}{n^2}\cdot\frac{n(n+1)}{2}\to\frac{1}{2}(n\to\infty).$$

不满足马尔可夫条件。

7.若 ξ_k 具有有限方差，服从同一分布，但各 k 间，ξ_k 和

ξ_{k+1} 有相关,而 ξ_k, ξ_l ($|k-l|\geqslant 2$) 是独立的,证明这时对 $\{\xi_k\}$大数定律成立。

证: 因为 ξ_k, ξ_l($|k-l|\geqslant 2$ 是独立的,所以

$$\frac{1}{n^2}D\left(\sum_{k=1}^{n}\xi_k\right)=\frac{1}{n^2}E\left[\sum_{k=1}^{n}(\xi_k-E\xi_k)\right]^2$$

$$=\frac{1}{n^2}\left[\sum_{k=1}^{n}E(\xi_k-E\xi_k)^2+\right.$$

$$\left.+2\sum_{k=1}^{n-1}E(\xi_k-E\xi_k)(\xi_{k+1}-E\xi_{k+1})\right]$$

$$=\frac{n}{n^2}D\xi_k+\frac{2}{n^2}\sum_{k=1}^{n-1}r_{k,k+1}\sigma^2\leqslant\frac{\sigma^2}{n}+\frac{2}{n}\sigma^2$$

$$=\frac{3}{n}\sigma^2\to 0(n\to\infty).$$

其中利用 $|r_{k,k+1}|\leqslant 1$且σ有限.马尔可夫条件成立,所以对序列$\{\xi_n\}$成立大数定律。

8.(伯恩斯坦定理)已知随机变量序列 ξ_1, ξ_2,…的方差有界,$D\xi_n\leqslant c$,并且当 $|i-j|\to\infty$时,相关系数 $r_{ij}\to 0$,证明对$\{\xi_n\}$成立大数定律。

证: 由题中条件可得,对任给$\varepsilon>0$,存在N,使当 $|j-i|>$ N时有 $|r_{ij}|<\frac{\varepsilon}{4c}$ (设$c>0$).则

$$\frac{1}{n^2}D\left(\sum_{k=1}^{n}\xi_k\right)=\frac{1}{n^2}\sum_{k=1}^{n}D\xi_k+$$

$$+\frac{1}{n^2}\cdot 2\sum_{n\geqslant j>i\geqslant 1}r_{ij}\sigma_i\sigma_j$$

$$\leqslant \frac{1}{n^2}\cdot nc+\frac{2}{n^2}\sum_{i<j}|r_{ij}|\sigma_i\sigma_j$$

$$\leqslant \frac{c}{n}+\frac{2}{n^2}\sum_{N\geqslant j-i,\ i<j}|r_{ij}|\sigma_i\sigma_j+$$

$$+\frac{2}{n^2}\sum_{j-i>N}|r_{ij}|\sigma_i\sigma_j.$$

在上式前一个和式中，i 可以依次取 $1,2,\cdots,n$；对每个固定的 i 来说，由于 $j-i\leqslant N$ 且 $i<j$，所以至多对应N项；从而和式中至多有nN项。在后一个和式中，由于 $j>i$，所以对 i 取 $1,2,\cdots,n$，至多依次对应 $n-1,n-2,\cdots 2,1$项，从而和式中至多有 $(n-1)+(n-2)+\cdots+2+1=\frac{1}{2}n(n-1)$ 项。利用 $|r_{ij}|\leqslant$ 可得

$$\frac{1}{n^2}D\left(\sum_{k=1}^{n}\xi_k\right)\leqslant\frac{c}{n}+\frac{2}{n^2}\cdot nN\cdot c+$$

$$+\frac{2}{n^2}\cdot\frac{n(n-1)}{2}\cdot\frac{\varepsilon}{4c}\cdot c$$

$$=\frac{c}{n}+\frac{2Nc}{n}+\left(1-\frac{1}{n}\right)\frac{\varepsilon}{4}.$$

当 n 充分大时，上式右方之值可以小于 ε，所以

$$\frac{1}{n^2}D\left(\sum_{k=1}^{n}\xi_k\right)\to 0(n\to\infty).$$

对$\{\xi_n\}$大数定律成立。

9．（格涅坚科定理）对随机变量序列$\{\xi_i\}$，若记

$\eta_n=\frac{1}{n}(\xi_1+\cdots+\xi_n)$，$a_n=\frac{1}{n}(E\xi_1+\cdots+E\xi_n)$，则$\{\xi_i\}$

服从大数定律的充要条件是

$$\lim_{n\to\infty} E\left\{\frac{(\eta_n-a_n)^2}{1+(\eta_n-a_n)^2}\right\}=0.$$

证：充分性。$s=t^2/(1+t^2)$是$t(>0)$的增函数，所以对任给$\varepsilon>0$有

$$P\{|\eta_n-a_n|\geqslant\varepsilon\}=\int_{|y-a_n|\geqslant\varepsilon} dF_{\eta_n}(y)\leqslant$$

$$\leqslant\frac{1}{\frac{\varepsilon^2}{1+\varepsilon^2}}\int_{|y-a_n|\geqslant\varepsilon}\frac{(y-a_n)^2}{1+(y-a_n)^2}dF_{\eta_n}(y)$$

$$\leqslant\frac{1+\varepsilon^2}{\varepsilon^2}E\left\{\frac{(\eta_n-a_n)^2}{1+(\eta_n-a_n)^2}\right\}.$$

所以当 $E\left\{\frac{(\eta_n-a_n)^2}{1+(\eta_n-a_n)^2}\right\}\to0$ 时有$P\{|\eta_n-a_n|\geqslant\varepsilon\}\to0$

此时$\{\xi_i\}$服从大数定律。

必要性。设$\{\xi_i\}$服从大数定律，即

$\lim\limits_{n\to\infty} P\{|\eta_n-a_n|\geqslant\varepsilon\}=0$，则对任给$\varepsilon>0$，存在$N$，当$n>N$时有$P\{|\eta_n-a_n|\geqslant\varepsilon\}\leqslant\varepsilon$。由$s=t^2/(1+t^2)$关于$t(>0)$的单调性和$0<s<1$得

$$0\leqslant E\left\{\frac{(\eta_n-a_n)^2}{1+(\eta_n-a_n)^2}\right\}\leqslant\frac{\varepsilon^2}{1+\varepsilon^2}P\{|\eta_n-a_n|<\varepsilon\}+$$

$$+1\cdot P\{|\eta_n-a_n|\geqslant\varepsilon\}<\varepsilon^2+\varepsilon<2\varepsilon(\text{当}\varepsilon<1\text{时}).$$

$$\therefore\quad \lim_{n\to\infty}E\left\{\frac{(\eta_n-a_n)^2}{1+(\eta_n-a_n)^2}\right\}=0.$$

10. 用斯特灵公式证明：当$n\to\infty$，$m\to\infty$，$n-m\to\infty$，而$m/n\to0$时，

$$\binom{2n}{n-m}\left(\frac{1}{2}\right)^{2n} \sim \frac{1}{\sqrt{n\pi}} e^{-m^2/n}$$

（原题中是$\frac{1}{\sqrt{n\pi}} e^{-2m^2/n}$）。

证：斯特灵公式为

$$m! = \sqrt{2\pi m}\, m^m e^{-m} e^{\theta_m}, \quad 0<\theta_m<\frac{1}{12m}.$$

由此得

$$\binom{2n}{n-m}\left(\frac{1}{2}\right)^{2n} = \frac{(2n)!}{(n-m)!\,(n+m)!}\left(\frac{1}{2}\right)^{2n}$$

$$= \frac{\sqrt{2\pi\cdot 2n}\,(2n)^{2n}e^{-2n}\left(\frac{1}{2}\right)^{2n}e^{\theta}}{\sqrt{2\pi(n-m)}\,(n-m)^{n-m}e^{-(n-m)}\sqrt{2\pi(n+m)}}\times$$

$$\times\frac{1}{(n+m)^{n+m}e^{-(n+m)}}$$

$$|\theta| < \frac{1}{12}\left(\frac{1}{2n}+\frac{1}{n-m}+\frac{1}{n+m}\right)$$

$$= \frac{1}{\sqrt{2\pi}}\sqrt{\frac{2n}{n^2-m^2}}\left(\frac{n}{n+m}\right)^{n+m}\left(\frac{n}{n-m}\right)^{n-m} e^{\theta}$$

$$= \frac{1}{\sqrt{2\pi}}\sqrt{\frac{2n}{n^2-m^2}}\left(1+\frac{m}{n}\right)^{-(n+m)}\left(1-\frac{m}{n}\right)^{-(n-m)} e^{\theta}$$

$$= \frac{1}{\sqrt{n\pi}}\frac{1}{\sqrt{1-\left(\frac{m}{n}\right)^2}}\frac{\left(1-\frac{m}{n}\right)^m}{\left(1-\frac{m^2}{n^2}\right)^n\left(1+\frac{m}{n}\right)^m} e^{\theta} \tag{1}$$

（若$\alpha\to 0$，$\beta\to\infty$，则当

$$\beta\alpha^2 = o(1) \tag{2}$$

时，才有下式成立：

$$(1+\alpha)^{\beta}=\left[(1+\alpha)^{1/\alpha}\right]^{\alpha\beta}\sim e^{\alpha\beta} \qquad (3)$$

此题未必满足(2)式，所以不加条件地利用(3)式证是不妥的。这里结论的证明很简单。若利用(3)式估计(1)式值，则应有

$$m^3/n^2=o(1),\quad m^4/n^3=o(1).$$

后一式蕴函在前一式中，即应补设前一条件成立，利用(3)才可证得结论。下面用另一种证法证明。）

视n，m为连续变量进行估值，然后再置n，m为取正整数的变量，结论也应成立。利用台劳展式

$$\ln(1+x)=\sum_{k=1}^{n}(-1)^{k-1}\frac{x^k}{n}+o(x^n)$$

$$(-1<x\leqslant 1),$$

由$m/n\to 0$得

$$\ln\frac{\left(1-\frac{m}{n}\right)^m}{\left(1-\frac{m^2}{n^2}\right)^n\left(1+\frac{m}{n}\right)^m}=m\ln\left(1-\frac{m}{n}\right)-$$

$$-n\ln\left(1-\frac{m^2}{n^2}\right)-m\ln\left(1+\frac{m}{n}\right)$$

$$=m\left(-\frac{m}{n}-\frac{m^2}{2n^2}-\frac{m^3}{3n^3}-\frac{m^4}{4n^4}-\frac{m^5}{5n^5}-\cdots\right)-$$

$$-n\left(-\frac{m^2}{n^2}-\frac{m^4}{2n^4}-\frac{m^6}{3n^6}\cdots\right)-$$

$$-m\left(\frac{m}{n}-\frac{m^2}{2n^2}+\frac{m^3}{3n^3}-\frac{m^4}{4n^4}+\frac{m^5}{5n^5}-\cdots\right)$$

$$=-\frac{m^2}{n}-\frac{1}{6}\frac{m^4}{n^3}-\frac{1}{15}\frac{m^6}{n^5}-\cdots$$

$$= -\frac{m^2}{n} - \frac{m^4}{n^3}\left(\frac{1}{6} + \frac{1}{15}\frac{m^2}{n^2} + \cdots\right)$$

$$= -\frac{m^2}{n} - \frac{m^4}{n^3}\left(\frac{1}{6} + o(1)\right),$$

$$\frac{\left(1 - \frac{m}{n}\right)^m}{\left(1 - \frac{m^2}{n^2}\right)^n\left(1 + \frac{m}{n}\right)^m} = e^{-\frac{m^2}{n} - \frac{m^4}{n^3}\left(\frac{1}{6} + o(1)\right)}$$

由题设条件得

$$\frac{m^4}{n^3} = \frac{m^2}{n}\left(\frac{m}{n}\right)^2 = o\left(\frac{m^2}{n}\right),$$

所以要证明的结论中只能是 $\frac{1}{\sqrt{n\pi}} e^{-m^2/n}$

在题设条件下显然有 $e^\theta \sqrt{1 - \left(\frac{m}{n}\right)^2} \to 1$，所以欲

$$\binom{2n}{n-m}\left(\frac{1}{2}\right)^{2n} \div \frac{1}{\sqrt{n\pi}} e^{-m^2/n}$$

$$= \frac{1}{\sqrt{n\pi}} \frac{1}{\sqrt{1 - \left(\frac{m}{n}\right)^2}} \frac{\left(1 - \frac{m}{n}\right)^m}{\left(1 - \frac{m^2}{n^2}\right)^n\left(1 + \frac{m}{n}\right)^m} e^\theta \div$$

$$\div \frac{1}{\sqrt{n\pi}} e^{-m/n^2}$$

$$= \frac{e^\theta}{\sqrt{1 - (m/n)^2}} e^{-\frac{m^4}{n^3}\left(\frac{1}{6} + o(1)\right)} \to 1,$$

必须且只需

$$\frac{m^4}{n^3}\left(\frac{1}{6}+o(1)\right)=o(1)\quad 即\frac{m^4}{n^3}\to 0.$$

这条件必须在题中补设出来，即再当$m^4/n^3\to 0$时有

$$\binom{2n}{n-m}\left(\frac{1}{2}\right)^{2n}\sim\frac{1}{\sqrt{n\pi}}e^{-m^2/n}$$

11.用(1.33)计算b(5 ;500,0.01)及b(40;10000,0.005),并与精确值比较。

解：(1.33)式为

$$P\{\mu_n=k\}=\frac{1}{\sqrt{npq}}\cdot\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x_k^2}\times$$

$$\times\left\{1+\frac{(x_k^3-3x_k)(q-p)}{6\sqrt{npq}}+o\left(\frac{1}{n}\right)\right\}$$

其中 $x_k=(k-np)/\sqrt{npq}$.

(1)b(5 ;500,0.01).$k=5, n=500, p=0.01, x_k=0$.

$$P\{\mu_n=k\}=\frac{1}{\sqrt{500\times0.01\times0.99}}\cdot\frac{1}{\sqrt{2\pi}}\cdot e^0$$
$$=0.17931.$$

准确值为0.17635.

(2)b(40;10000,0.005).$k=40, n=10000, p=0.005$.

$$x_k=\frac{40-50}{\sqrt{10000\times0.005\times0.995}}=-1.417762$$

$$e^{-\frac{1}{2}x_k^2}=0.366035$$

$$1+\frac{(x_k^3-3x_k)(q-p)}{6\sqrt{npq}}=0.1032832$$

$$\frac{1}{\sqrt{npq\cdot 2\pi}}=\frac{1}{\sqrt{10000\times 0.005\times 0.995\times 2\pi}}$$

$$=0.366035$$

$$P\{\mu_n=k\}=0.0565605\times 1.0328323\times 0.366035\approx$$

$$\approx 0.0214.$$

准确值为0.0214。

12．某计算机系统有120个终端，每个终端有5％时间在使用，若各个终端使用与否是相互独立的，试求有10个或更多终端在使用的概率

解：每个终端在某时刻使用的概率为0.05，ξ表示在某时刻同时使用的终端数．则

$$P\{\xi=k\}=C_{120}^{k}(0.05)^{k}(0.95)^{120-k}$$

由积分极限定理得

$$P\{120\geqslant\xi\geqslant 10\}=1-P\{\xi<10\}$$

$$=1-P\left\{\frac{\xi-6}{\sqrt{120\times 0.05\times 0.95}}<\frac{10-6}{\sqrt{120\times 0.05\times 0.95}}\right\}$$

$$\approx 1-\Phi(1.675)=0.047.$$

即有10个或更多个终端在使用的概率为0.047。

13．求证，在$x>0$时有不等式

$$\frac{x}{1+x^2}e^{-\frac{1}{2}x^2}\leqslant\int_x^{\infty}e^{-\frac{1}{2}t^2}dt\leqslant\frac{1}{x}e^{-\frac{1}{2}x^2}$$

证：当$x>0$时有

$$\int_x^{\infty}e^{-\frac{1}{2}x^2}dt\leqslant\int_x^{\infty}\frac{t}{x}e^{-\frac{1}{2}t^2}dt$$

$$=\frac{1}{x}\left(-e^{-\frac{1}{2}t^2}\right)\Big|_x^{\infty}=\frac{1}{x}e^{-\frac{1}{2}x^2}$$

$$\int_x^{\infty} e^{-\frac{1}{2}t^2} dt \geqslant \int_x^{\infty} \frac{t^4+2t^2-1}{t^4+2t^2+1} e^{-\frac{1}{2}t^2} dt$$

$$= \int_x^{\infty} d\left(-\frac{t}{1+t^2} e^{-\frac{1}{2}t^2}\right)$$

$$= \frac{-t}{1+t^2} e^{-\frac{1}{2}t^2} \Big|_x^{\infty} = \frac{x}{1+x^2} e^{-\frac{1}{2}x^2}$$

所以不等式成立。

14. 用德莫哇佛—拉普拉斯定理证明，在贝努里试验中，$0<p<1$，则不管 k 是如何大的常数，总有

$$P\{|\mu_n - np| < k\} \to 0 (n\to\infty).$$

证：利用德莫哇佛—拉普拉斯积分定理得

$$P\{|\mu_n - np| < k\} = P\left\{\frac{|\mu_n - np|}{\sqrt{npq}} < \frac{k}{\sqrt{npq}}\right\}$$

$$\approx \frac{1}{\sqrt{2\pi}} \int_{-k/\sqrt{npq}}^{k/\sqrt{npq}} e^{-\frac{1}{2}x^2} dx$$

在如上积分中，积分区间长度 $2k/\sqrt{npq} \to 0$，所以

$$P\{|\mu_n - np| < k\} \to 0 \quad (n\to\infty).$$

15. 用车贝晓夫不等式确定当掷一均匀铜币时，需投多少次，才能保证使得正面出现的频率在0.4至0.6之间的概率不小于90%. 并用正态逼近计算同一问题。

解：设需要投掷 n 次，用车贝晓夫不等式得($p=0.5$)

$$P\left\{0.4 < \frac{\mu_n}{n} < 0.6\right\} = P\left\{\left|\frac{\mu_n}{n} - 0.5\right| < 0.1\right\}$$

$$\geqslant 1 - \frac{1}{0.1^2} \cdot \frac{1}{n^2} \cdot n \cdot \frac{1}{2} \cdot \frac{1}{2} = 0.9$$

$$\frac{1}{0.04n}=0.1，取n\geqslant\frac{1000}{4}=250.$$

用积分极限定理得

$$P\left\{0.4<\frac{\mu_n}{n}<0.6\right\}=P\left\{\frac{|\mu_n-0.5|}{\sqrt{npq}}<\frac{0.1\sqrt{n}}{\sqrt{pq}}\right\}$$

$$\approx\Phi\left(\frac{1}{5}\sqrt{n}\right)-\Phi\left(-\frac{1}{5}\sqrt{n}\right)=2\Phi\left(\frac{1}{5}\sqrt{n}\right)-1=$$

$$=0.9$$

$\Phi\left(\frac{1}{5}\sqrt{n}\right)=0.95$，$\frac{1}{5}\sqrt{n}=1.645$，$n=67.65$ 取 $n\geqslant68$.

16.用车贝晓夫不等式及德莫哇佛—拉普拉斯极限定理估计下面概率：

$$P\left\{\left|\frac{\mu_n}{n}-p\right|\geqslant\varepsilon\right\}$$

并进行比较.这里 μ_n是 n 次贝努里试验中成功总次数，p为每次成功的概率.

解：利用车贝晓夫不等式估计值为：

$$P\left\{\left|\frac{\mu_n}{n}-p\right|\geqslant\varepsilon\right\}\leqslant\frac{npq/n^2}{\varepsilon^2}=$$

$$=\frac{pq}{n\varepsilon^2}.$$

利用德莫哇佛—拉普拉斯极限定理估值为：

$$P\left\{\left|\frac{\mu_n}{n}-p\right|\geqslant\varepsilon\right\}=P\left\{\frac{|\mu_n-np|}{\sqrt{npq}}\geqslant\frac{\varepsilon\sqrt{n}}{\sqrt{pq}}\right\}$$

$$\approx\frac{2}{\sqrt{2\pi}}\int_{\varepsilon\sqrt{n}/\sqrt{pq}}^{\infty}e^{-\frac{1}{2}t^2}dt$$

$$<\sqrt{\frac{2}{\pi}}\frac{\sqrt{pq}}{\varepsilon\sqrt{n}}\int_{\varepsilon\sqrt{n}/\sqrt{pq}}^{\infty} te^{-\frac{1}{2}t^2}\,dt$$

$$=\frac{1}{\varepsilon}\sqrt{\frac{2pq}{n\pi}}\,e^{-\frac{\varepsilon^2 n}{2pq}}=o\left(\frac{pq}{n\varepsilon^2}\right)\ (n\to\infty)$$

两者比较，后者估计精确得多。

17. 现有一大批种子，其中良种占$\frac{1}{6}$，今在其中任选6000粒，试问在这些种子中，良种所占的比例与$\frac{1}{6}$之差小于1%的的概率是多少？

解： 任选6000粒可看作6000重贝努里试验，$p=\frac{1}{6}$，由积分极限定理得

$$P\left\{\left|\frac{\mu}{6000}-\frac{1}{6}\right|<0.01\right\}=$$

$$=P\left\{\frac{|\mu-6000\times\frac{1}{6}|}{\sqrt{6000\times\frac{1}{6}\times\frac{5}{6}}}\leqslant\frac{0.01\sqrt{6000}}{\sqrt{\frac{1}{6}\times\frac{5}{6}}}\right\}$$

$$\approx\Phi(2.078)-\Phi(-2.078)=2\Phi(2.078)-1$$

$$=2.98124-1\approx0.96.$$

18. 种子中良种占$\frac{1}{6}$，我们有99%的把握断定，在6000粒种子中良种所占的比例与$\frac{1}{6}$之差是多少？这时相应的良种数落在哪个范围内？

解： 与上题同理得

$$P\left\{\left|\frac{\mu}{6000}-\frac{1}{6}\right|<\varepsilon\right\}=$$

$$=P\left\{\frac{|\mu-6000\times\frac{1}{6}|}{\sqrt{6000\times\frac{1}{6}\times\frac{5}{6}}}\leqslant120\sqrt{3}\,\varepsilon\right\}=0.99,$$

$2\Phi(120\sqrt{3}\,\varepsilon)-1=0.99$，$\Phi(120\sqrt{3}\,\varepsilon)=0.995$，

$120\sqrt{3}\,\varepsilon=2.58$，$\varepsilon=0.0124$.

把$\varepsilon=0.0124$代入上式计算得

$$P\left\{\left|\frac{\mu}{6000}-\frac{1}{6}\right|<0.0124\right\}=P\left\{\left|\mu-1000\right|<74.4\right\}$$

$$=P\{925<\mu<1075\}=0.99.$$

所以相应的良种数应落在925粒与1075粒之间。

19. 蒲丰试验中掷铜币4040次，出正面2048次，试计算当重复蒲丰试验时，正面出现的频率与概率之差的偏离程度，不大于蒲丰试验中所发生的偏差的概率。

解：在蒲丰试验中，频率与概率之差为

$$\frac{2048}{4040}-\frac{1}{2}=\frac{28}{4040}=0.00693.$$

由积分极限定理得要求的概率为

$$P\left\{\left|\frac{\mu}{4040}-\frac{1}{2}\right|\leqslant 0.00693\right\}=$$

$$=P\left\{\frac{|\mu-2020|}{\sqrt{4040\times\frac{1}{4}}}\leqslant\frac{0.00693\times\sqrt{4040}}{1/2}\right\}$$

$$=P\left\{\frac{|\mu-2020|}{\sqrt{4040\times\frac{1}{4}}}\leqslant 0.88\right\}\approx 2\Phi(0.88)-1$$

$$=2\times 0.8106-1\approx 0.621.$$

20. 设分布函数列$\{F_n(x)\}$弱收敛于连续的分布函数$F(x)$，试证这收敛对$x\in R^1$是一致的。

证：由于$F(x)$有界非降，$F(-\infty)=0$，$F(+\infty)=1$，故对任意$\varepsilon>0$，可找到$M>0$，使当$x\geqslant M$时有

$$1-F(x)<\varepsilon, \tag{1}$$

且当$x\leqslant -M$时有

$$F(x)<\varepsilon. \tag{2}$$

由于$F_n(x)$处处收敛于$F(x)$，故存在一正整数N_1，使当

$n>N_1$时，一方面有

$$|F_n(-M)-F(-M)|<\varepsilon.$$

由(2)得

$$F_n(-M)<2\varepsilon; \tag{3}$$

另一方面又有

$$|F_n(M)-F(M)|<\varepsilon,$$

由(1)得

$$1-F_n(M)<2\varepsilon, \tag{4}$$

因此，对$x<-M$，若$n\geqslant N_1$，则由(2)，(3)有

$$\begin{aligned}|F_n(x)-F(x)| &< |F_n(x)|+|F(x)|\\ &\leqslant F_n(-M)+F(-M)<3\varepsilon.\end{aligned} \tag{5}$$

同样，对$x>M$，如果$n>N_1$，则由(1)，(4)有

$$\begin{aligned}|F_n(x)-F(x)| &= |(1-F_n(x))-(1-F(x))|\\ &< |1-F_n(x)|+|1-F(x)|\\ &\leqslant 1-F_n(M)+1-F(M)<3\varepsilon.\end{aligned} \tag{6}$$

在有限闭区间$[-M,M]$上，$F(x)$连续，故也均匀连续，因而在$[-M,M]$上可找到t个点$x_1,x_2,\cdots,x_t$，$x_1=-M$，$x_t=M$，使

$$F(x_{i+1})-F(x_i)<\varepsilon\ (i=1,2,\cdots,t-1). \tag{7}$$

还可找到$N_2>N_1$，使在此t个点中的每一点上，当$n>N_2$时有

$$|F_n(x_i)-F(x_i)|<\varepsilon. \tag{8}$$

于$[-M,M]$中任取一x，则此x必属于某一$[x_i,x_{i+1}]$。因此当$n>N_2$时，由(8)得

$$F_n(x)\leqslant F_n(x_{i+1})<F(x_{i+1})+\varepsilon, \tag{9}$$

及
$$F_n(x)\geqslant F_n(x_i)>F(x_i)-\varepsilon. \tag{10}$$

由此及(9)，(7)得

$$F_n(x)-F(x)<F(x_{i+1})-F(x)+\varepsilon$$
$$\leqslant F(x_{i+1})-F(x_i)+\varepsilon<2\varepsilon. \tag{11}$$

同样由(10)及(7)得

$$F_n(x)-F(x)>F(x_i)-F(x)-\varepsilon$$
$$\geqslant F(x_i)-F(x_{i+1})-\varepsilon>-2\varepsilon. \tag{12}$$

故当$N>N_2$时，由(5),(6),(11),(12)得,对任意$x\in R_1$有

$$|F_n(x)-F(x)|<3\varepsilon.$$

21.设$F_n(x)$为一列正态分布函数,收敛于分布函数$F(x)$试证$F(x)$也是正态分布函数。

证：题中收　系指弱收敛。

设F_n,F的特征函数分别为f_n及f,由假定

$$f(t)=\lim_{n\to\infty}f_n(t)=\lim_{n\to\infty}\exp\{ia_nt-\tfrac{1}{2}\sigma_n^2\ t^2\} \tag{1}$$

因此若能证明极限

$$a=\lim_{n\to\infty}a_n,\quad \sigma^2=\lim_{n\to\infty}\sigma_n^2$$

存在,则本题得以证明。由(1)得

$$\exp\{-\tfrac{1}{2}\sigma_n^2\ t^2\}=|f_n(t)|\to|f(t)|.$$

因f(t)连续而且f(0)=1,故必存在$t_0\neq0$使$f(t_0)\neq0$,从而

$$-\tfrac{1}{2}\sigma_n^2\ t_0^2\to\log|f(t_0)|\neq-\infty$$

故得证极限$\sigma^2=\lim\limits_{n\to\infty}\sigma_n^2$存在。

由上一段证明及$F_n\xrightarrow{W}F$知,关于$t\in[0,1]$一致地有

$\exp\{\frac{1}{2}\sigma_n^2\ t^2\}\to\exp\{\frac{1}{2}\sigma^2t^2\}$, $f_n(t)\to f(t)$,因而

$$\exp\{ia_n t\}=f_n(t)\exp\{\tfrac{1}{2}\ \sigma_n^2\ t^2\}\to f(t)\exp\{\tfrac{1}{2}\sigma^2 t^2\} \tag{2}$$

故得

$$|f(t)\exp\{\tfrac{1}{2}\sigma^2 t^2\}|=\lim_{n\to\infty}|\exp\{ia_n t\}|=1.$$

今取积分路线

C_n: $z=\exp\{ia_n t\}$, $0\leqslant t\leqslant 1$,

C: $z=f(t)\exp\{\tfrac{1}{2}\sigma^2 t^2\}$, $0\leqslant t\leqslant 1$;

由(2)知$C_n\to C$,(即指表示C_n的函数$\exp\{ia_n t\}$,在$0\leqslant t\leqslant 1$上一致收敛于表示C的函数$f(t)\exp\{\tfrac{1}{2}\sigma^2 t^2\}$)。既然在积分路上$|z|=1\neq 0$,故

$$ia_n=\int_{C_n}\frac{dz}{z}\to\int_C\frac{dz}{z}.$$

即极限$a=\lim\limits_{n\to\infty}a_n$存在。

若$\sigma^2=0$,则极限分布是退化分布;若$\sigma^2\neq 0$,则极限分布服从$N(a,\sigma^2)$。

22.试证若正态随机变量序列依概率收敛,则其数学期望及方差也收敛。

证: 由$X_n\xrightarrow{P}X$可推得$X_n\xrightarrow{L}X$,从而$F_n(x)\xrightarrow{W}F(x)$,由上题即得证。

23.若X_n为多维正态随机向量,$X_n\xrightarrow{P}X$,试证X为正态向量。

证: 利用教本P222定理6(见第四章第57题解答).设

$X_n=(x_1^{(n)},\cdots,x_k^{(n)})$,$X=(x_1^{(0)},\cdots,x_k^{(0)})$,由$X_n\xrightarrow{P}X$

得,任一线性组合

$$Y_n=\sum_{j=1}^{k} l_j X_j^{(n)} \xrightarrow{P} Y=\sum_{j=1}^{k} l_j X_j^{(0)}$$

（这可用26，27题证法证明），由此可得 $Y_n \xrightarrow{L} Y$，从而$F_{y_n}(x)$ $\xrightarrow{W}$ $F_Y(x)$。由上述定理 6 知Y_n服从一维正态分布，由21题结论得Y也服从一维正态分布，再利用上述定理 6 得，X为多维正态随机向量。

24．若X_n的概率分布为$\begin{pmatrix} 0 & n \\ 1-\frac{1}{n} & \frac{1}{n} \end{pmatrix}$，试证相应的分布函数收敛，但矩不收敛。

证：

$$F_n(x)=P\{X_n<x\}=\begin{cases} 0, & \text{当}x\leqslant 0 \\ 1-\frac{1}{n}, & \text{当}0<x\leqslant n \\ 1, & \text{当}x>n. \end{cases}$$

令$n\to\infty$得

$$F_n(x)\to F(x)=\begin{cases} 0, & x\leqslant 0 \\ 1, & x>0. \end{cases}$$

这说明分布函数收敛。但

$$EX_n=1,\ EX=0,\ EX_n\nrightarrow EX(n\to\infty).$$

当$k>1$时，

$$EX_n^k=n^k\cdot\frac{1}{n}=n^{k-1},$$

$$E(X_n-EX_n)^k=E(X_n-1)^k=$$

$$=(-1)^k(1-\frac{1}{n})+(n-1)^k\cdot\frac{1}{n}$$

所以当 $n\to\infty$时，$EX_n^k\to\infty$，$E(X_n-EX_n)^k\to\infty$。由此知其中心矩，原点矩均不收敛。

25. 随机变量序列$\{\xi_n\}$具有分布函数$\{F_n(x)\}$，且$F_n(x)\xrightarrow{w} F(x)$，又$\{\eta_n\}$依概率收敛于常数$c>0$。试证：
（Ⅰ）$\zeta_n=\xi_n+\eta_n$的分布函数收敛于$F(x-c)$；（Ⅱ）$\zeta_n=\xi_n/\eta_n$的分布函数收敛于$F(cx)$。

证： 题中分布函数收敛系指弱收敛。

（Ⅰ）设$x-c$是$F(x)$的连续点，现证$F_{\zeta_n}(x)\to F(x-c)$。对任给$\varepsilon>0$，有

$$\{\xi_n+\eta_n<x\}=\{\xi_n+\eta_n<x,|\eta_n-c|<\varepsilon\}\cup\{\xi_n+\eta_n<x,|\eta_n-c|\geqslant\varepsilon\}.$$

上式中右边两事件依次记为S_1,S_2，则$S_1\cap S_2=\phi$，

$$F_{\zeta_n}(x)=P\{\xi_n+\eta_n<x\}=P(S_1)+P(S_2). \tag{1}$$

我们有

$$\begin{aligned}&P\{\xi_n<x-(c+\varepsilon)\}-P\{|\eta_n-c|\geqslant\varepsilon\}\\&\quad\leqslant P\{\xi_n<x-(c+\varepsilon),|\eta_n-c|\leqslant\varepsilon\}\leqslant P(S_1)\\&\quad\leqslant P\{\xi_n<x-(c-\varepsilon),|\eta_n-c|<\varepsilon\}\\&\quad\leqslant P\{\xi_n<x-(c-\varepsilon)\}.\end{aligned} \tag{2}$$

由(1)，(2)得

$$\begin{aligned}&P\{\xi_n<x-(c+\varepsilon)\}-P\{|\eta_n-c|\geqslant\varepsilon\}+P(S_2)\leqslant\\&\leqslant P\{\xi_n+\eta_n<x\}\leqslant P\{\xi_n<x-(c-\varepsilon)\}+P(S_2).\end{aligned}$$

此式对任意n成立，所以

$$\underline{\lim_{n\to\infty}}P\{\xi_n<x-(c+\varepsilon)\}+\underline{\lim_{n\to\infty}}[P(S_2)-P\{|\eta_n-c|\geqslant\varepsilon\}]$$

$$\leqslant\underline{\lim_{n\to\infty}}F_{\zeta_n}(x)\leqslant\overline{\lim_{n\to\infty}}F_{\zeta_n}(x)$$

$$\leqslant\overline{\lim_{n\to\infty}}P\{\xi_n<x-c+\varepsilon\}+\overline{\lim_{n\to\infty}}P(S_2) \tag{3}$$

由$\eta_n \xrightarrow{P} c$得

$$P(S_2) \leqslant P\{|\eta_n - c| \geqslant \varepsilon\} \to 0 (n \to \infty).$$

再适当选取 ε 使$x-c-\varepsilon$, $x-c+\varepsilon$同是$F(x)$的连续点,利用弱收敛性由(3)可得

$$F(x-c-\varepsilon) \leqslant \varliminf_{n\to\infty} F_{\zeta_n}(x) \leqslant \varlimsup_{n\to\infty} F_{\zeta_n}(x)$$

$$\leqslant F(x-c+\varepsilon). \tag{4}$$

由于$F(x)$单调增加,其至多有可列个不连续点,这里对 ε 的限制丝毫不影响以下结论成立.由于 ε 是任意的且$x-c$是$F(x)$的连续点,由(4)得

$$\lim_{n\to\infty} F_{\zeta_n}(x) = F(x-c).$$

所以 $F_{\zeta_n}(x) \xrightarrow{W} F(x-c)$.

(Ⅱ)设$cx(x\neq 0)$是 $F(x)$的连续点.对任给$\varepsilon>0$ $(0<\varepsilon<c)$,

$$\{\xi_n/\eta_n<x\} = \{\xi_n/\eta_n<x, |\eta_n-c|<\varepsilon\} \cup$$
$$\cup \{\xi_n/\eta_n<x, |\eta_n-c| \geqslant \varepsilon\} = S_1 \cup S_2 (\text{记}),$$

则 $S_1 \cap S_2 = \phi, P(S_2) \leqslant P\{|\eta_n - c| \geqslant \varepsilon\} \to 0 (n\to\infty)$.

另外, $P(S_1)$介于如下两概率之间:

$$P\{\xi_n<(c-\varepsilon)x, |\eta_n-c|<\varepsilon\},$$
$$P\{\xi_n<(c+\varepsilon)x, |\eta_n-c|<\varepsilon\}.$$

对这两个概率值又分别有

$$0 \leqslant P\{\xi_n<(c-\varepsilon)x\} - P\{\xi_n<(c-\varepsilon)x, |\eta_n-c|<\varepsilon\}$$
$$\leqslant P\{|\eta_n-c| \geqslant \varepsilon\},$$
$$0 \leqslant P\{\xi_n<(c+\varepsilon)x\} - P\{\xi_n<(c+\varepsilon)x, |\eta_n-c|<\varepsilon\}$$
$$\leqslant P\{|\eta_n-c| \geqslant \varepsilon\}.$$

取极限可得，当$x>0$时有(若$x<0$,则下式前后两项分别改成取上,下极限,且调换前后之位置)，

$$\varliminf_{n\to\infty} F_n((c-\varepsilon)x)\leqslant \varliminf_{n\to\infty} F_{\zeta_n}(x)$$

$$\leqslant \varlimsup_{n\to\infty} F_{\zeta_n}(x)\leqslant \varlimsup_{n\to\infty} F_n((c+\varepsilon)x).$$

可适取ε,使$(c-\varepsilon)x$与$(c+\varepsilon)x$都是$F(x)$的连续点.当$x>0$时.由弱收敛性得(若$x<0$，则前后两项调换位置)

$$F((c-\varepsilon)x)\leqslant\varliminf_{n\to\infty} F_{\zeta_n}(x)\leqslant \varlimsup_{n\to\infty} F_{\zeta_n}(x)\leqslant$$

$$\leqslant F((c+\varepsilon)x).$$

由ε的任意性及cx是$F(x)$的连续点得

$$\lim_{n\to\infty} F_{\zeta_n}(x)=F(cx).$$

若$cx=0$(从而$x=0$)是$F(x)$的连续点，则对任意$\varepsilon>0$ $(\varepsilon<c)$有

$$F_{\zeta_n}(0)=p\{\xi_n/\eta_n<0\}$$

$$=P\{\xi_n/\eta_n<0,\ |\eta_n-c|<\varepsilon\}+P\{\xi_n/\eta_n<0,\ |\eta_n-c|\geqslant\varepsilon\}$$

$$=P\{\xi_n<0,\ |\eta_n-c|<\varepsilon\}+P\{\xi_n/\eta_n<0,\ |\eta_n-c|\geqslant\varepsilon\}$$

$$=P\{\xi_n<0\}-P\{\xi_n<0,\ |\eta_n-c|\geqslant\varepsilon\}+$$

$$+P\{\xi_n/\eta_n<0,\ |\eta_n-c|\geqslant\varepsilon\}.$$

等式右边三项中，由$F_n(x)\xrightarrow{W}F(x)$得第一项$P\{\xi_n<0\}=F_n(0)\to F(0)$,其余两项中概率值均不超过$P\{|\eta_n-c|\geqslant\varepsilon\}$,所以右边从而左边极限存在.取极限可得$\lim\limits_{n\to\infty} F_{\eta_n}(0)=F(0)$.

至此得证$F_{\zeta_n}(x)\xrightarrow{W}F(cx)$.

26.试证

(1) $X_n \xrightarrow{P} X \Rightarrow X_n - X \xrightarrow{P} 0$;

(2) $X_n \xrightarrow{P} X$, $X_n \xrightarrow{P} Y \Rightarrow P\{X = Y\} = 1$;

(3) $X_n \xrightarrow{P} X \Rightarrow X_n - X_m \xrightarrow{P} 0 \ (n, m \to \infty)$;

(4) $X_n \xrightarrow{P} X$, $Y_n \xrightarrow{P} Y \Rightarrow X_n \pm Y_n \xrightarrow{P} X \pm Y$;

(5) $X_n \xrightarrow{P} X$, k是常数$\Rightarrow kX_n \xrightarrow{P} kX$;

(6) $X_n \xrightarrow{P} X \Rightarrow X_n^2 \xrightarrow{P} X^2$;

(7) $X_n \xrightarrow{P} a$, $Y_n \xrightarrow{P} b$, a,b 常数$\Rightarrow X_n Y_n \xrightarrow{P} ab$;

(8) $X_n \xrightarrow{P} 1 \Rightarrow X_n^{-1} \xrightarrow{P} 1$;

(9) $X_n \xrightarrow{P} a$, $Y_n \xrightarrow{P} b$, a,b 常 数, $b \neq 0 \Rightarrow X_n Y_n^{-1} \xrightarrow{P} ab^{-1}$;

(10) $X_n \xrightarrow{P} X$, Y是随机变量$\Rightarrow X_n Y \xrightarrow{P} XY$.

(11) $X_n \xrightarrow{P} X$, $Y_n \xrightarrow{P} Y \Rightarrow X_n Y_n \xrightarrow{P} XY$.

证: (1)$P\{|X_n - X - 0| \geqslant \varepsilon\} = P\{|X_n - X| \geqslant \varepsilon\} \to 0$,

$$\therefore \quad X_n - X \xrightarrow{P} 0 \ (n \to \infty),$$

(2)对任给$\varepsilon > 0$,

$$P\{|X - Y| \geqslant \varepsilon\} = P\{|X - X_n + X_n - Y| \geqslant \varepsilon\}$$
$$\leqslant P\{|X_n - X| \geqslant \tfrac{1}{2}\varepsilon\} + P\{|X_n - Y| \geqslant \tfrac{1}{2}\varepsilon\} \to 0 \quad (n \to \infty).$$

由ε的任意性得$P\{X \neq Y\} = 0$，所以$P\{X = Y\} = 1$.

(3)$P\{|X_n - X_m| \geqslant \varepsilon\} = P\{|X_n - X + X - X_m| \geqslant \varepsilon\}$
$$\leqslant P\{|X_n - X| \geqslant \tfrac{1}{2}\varepsilon\} + P\{|X_m - X| \geqslant \tfrac{1}{2}\varepsilon\} \to 0,$$

$\therefore\quad X_n - X_m \xrightarrow{P} 0\ (n, m\to\infty)$.

(4) $P\{|(X_n \pm Y) - (X \pm Y_n)| \geqslant \varepsilon\} =$

$= P\{|(X_n - X) \pm (Y_n - Y)| \geqslant \varepsilon\} \leqslant$

$\leqslant P\{|X_n - X| \geqslant \frac{1}{2}\varepsilon\} + P\{|Y_n - Y| \geqslant \frac{1}{2}\varepsilon\} \to 0$,

$\therefore\quad X_n \pm Y_n \xrightarrow{P} X \pm Y\ (n\to\infty)$

(5) 若$k=0$，显然有$kX_n \xrightarrow{P} kX(n\to\infty)$，若$k\neq 0$，则

$P\{|kX_n - kX| \geqslant \varepsilon\} = P\{|k|\ |X_n - X| \geqslant \varepsilon\}$

$= P\{|X_n - X| \geqslant \varepsilon/|k|\} \to 0$,

$\therefore\quad kX_n \xrightarrow{P} kX(n\to\infty)$.

(6) $P\{|X_n^2 - X^2| \geqslant \varepsilon\} =$

$= P\{|(X_n - X)^2 + 2X(X_n - X)| \geqslant \varepsilon\}$

$\leqslant P\{|X_n - X|^2 \geqslant \frac{1}{2}\varepsilon\} + P\{2|X|\ |X_n - X| \geqslant \frac{1}{2}\varepsilon\}$

$= P\{|X_n - X| \geqslant \sqrt{\frac{1}{2}\varepsilon}\} + P\{|X|\ |X_n - X| \geqslant \frac{1}{4}\varepsilon\}$

对任给$\delta>0$，取$M>0$，使$P\{|X|>M\}<\frac{1}{3}\delta$，再取$N$使当$n>N$时有$P\{|X_n - X| \geqslant \varepsilon/(4M)\}<\frac{1}{3}\delta$，且$P\{|X_n - X| \geqslant \sqrt{\frac{1}{2}\varepsilon}\}<\frac{\delta}{3}$

因为

$\{|X|\ |X_n - X| \geqslant \frac{1}{4}\varepsilon\} \subset$

$\{|X|>M\} \cup \{|X| \leqslant M,\ |X|\ |X_n - X| \geqslant \frac{1}{4}\varepsilon\}$

$\subset \{|X|>M\} \cup \{|X_n - X| \geqslant \varepsilon/(4M)\}$

所以当$n>N$时有

$P\{|X_n^2 - X^2| \geqslant \varepsilon\} \leqslant P\{|X_n - X| \geqslant \sqrt{\frac{1}{2}\varepsilon}\} +$

$+ P\{|X|>M\} + P\{|X_n - X| \leqslant \varepsilon/(4M)\}$

$$\leqslant \frac{1}{3}\delta + \frac{1}{3}\delta + \frac{1}{3}\delta = \delta,$$

从而 $\lim\limits_{n\to\infty} P\{|X_n^2 - X^2| \geqslant \varepsilon\} = 0$ 即 $X_n^2 \xrightarrow{P} X^2 (n\to\infty)$，

(7) $P\{|X_nY_n - ab| \geqslant \varepsilon\} =$

$$= P\{|X_nY_n - aY_n - bX_n + ab + aY_n + bX_n - 2ab| \geqslant \varepsilon\}$$

$$= P\{|(X_n - a)(Y_n - b) + a(Y_n - b) + b(X_n - a)| \geqslant \varepsilon\}$$

$$\leqslant P\{|(X_n - a)(Y_n - b)| \geqslant \frac{1}{3}\varepsilon\} +$$

$$+ P\{|a(Y_n - b)| \geqslant \frac{1}{3}\varepsilon\} + P\{|b(X_n - a)| \geqslant \frac{1}{3}\varepsilon\}$$

$$\leqslant P\{|X_n - a| \geqslant \sqrt{\tfrac{1}{3}\varepsilon}\} + P\{|Y_n - b| \geqslant \sqrt{\tfrac{1}{3}\varepsilon}\}$$

$$+ P\{|a|\ |Y_n - b| \geqslant \frac{1}{3}\varepsilon\} + P\{|b|\ |X_n - a| \geqslant \frac{1}{3}\varepsilon\}$$

$$\to 0,$$

$\therefore\quad X_nY_n \xrightarrow{P} ab (n\to\infty)$.

(8) $P\{|X_n^{-1} - 1| \geqslant \varepsilon\} = P\{|1 - X_n| / |X_n| \geqslant \varepsilon\}$

$$\leqslant P\{|X_n| \leqslant \frac{1}{2}\} + P\{|X_n| > \frac{1}{2}, |1 - X_n| / |X_n| \geqslant \varepsilon\}$$

$$\leqslant P\{|X_n - 1| \geqslant \frac{1}{2}\} + P\{|1 - X_n| \geqslant \frac{1}{2}\varepsilon\} \to 0,$$

$\therefore\quad X_n^{-1} \xrightarrow{P} 1 (n\to\infty)$.

(9)在(8)中令$X_n = Y_n/b$，再利用(5)由 $Y_n \xrightarrow{P} b$可证得$Y_n^{-1} \xrightarrow{P} b^{-1}$，再视(7)中$Y_n$为这里$Y_n^{-1}$即得证。

(10)对任给$\delta>0$，取$M>0$，使$P\{|Y|>M\}<\frac{1}{2}\delta$。再取N，使当$n>N$时，$P\{|X_n - X| \geqslant \varepsilon/M\} < \frac{1}{2}\delta$，则

$$P\{|X_nY - XY| \geqslant \varepsilon\} = P\{|Y|\ |X_n - X| \geqslant \varepsilon\}$$

$$\leqslant P\{|Y| > M\} + P\{|Y| \leqslant M, |Y|\ |X_n - X| \geqslant \varepsilon\}$$

$$\leqslant P\{|Y| > M\} + P\{|X_n - X| \geqslant \varepsilon/M\} < \frac{1}{2}\delta + \frac{1}{2}\delta = \delta,$$

$\therefore \quad X_nY \xrightarrow{P} XY\ (n\to\infty)$.

(11)对任给$\delta>0$，取$M>0$及N，使当$n>N$时如下五式同时成立：$P\{|X|\geqslant\frac{1}{2}M\}<\frac{1}{8}\delta$，$P\{|X_n-X|\geqslant\frac{1}{2}M\}<\frac{1}{8}\delta$，$P\{|Y|\geqslant M\}<\frac{1}{4}\delta$，$P\{|Y_n-Y|\geqslant\varepsilon/(2M)\}<\frac{1}{4}\delta$，$P\{|X_n-X|\geqslant\varepsilon/(2M)\}<\frac{1}{4}\delta$。则当$n>N$时有

$$
\begin{aligned}
P\{|X_n|\geqslant M\}&=P\{|X|\geqslant\tfrac{1}{2}M,\ |X_n|\geqslant M\}+\\
&+P\{|X|<\tfrac{1}{2}M,\ |X_n|\geqslant M\}\\
&\leqslant P\{|X|\geqslant\tfrac{1}{2}M\}+P\{|X|<\tfrac{1}{2}M,\ |X_n-X|\geqslant\tfrac{1}{2}M\}\\
&<\tfrac{1}{8}\delta+\tfrac{1}{8}\delta=\tfrac{1}{4}\delta,
\end{aligned}
$$

从而

$$
\begin{aligned}
&P\{|X_nY_n-XY|\geqslant\varepsilon\}=\\
&=P\{X_nY_n-X_nY+X_nY-XY|\geqslant\varepsilon\}\leqslant\\
&\leq P\{|X_n|\ |Y_n-Y|\geqslant\tfrac{1}{2}\varepsilon\}+P\{|Y|\ |X_n-X|\geqslant\tfrac{1}{2}\varepsilon\}\\
&\leqslant P\{|X_n|\geqslant M\}+P\{|X_n|<M,|X_n|\ |Y_n-Y|\geqslant\tfrac{1}{2}\varepsilon\}\\
&+P\{|Y|\geqslant M\}+P\{|Y|<M,|Y|\ |X_n-X|\geqslant\tfrac{1}{2}\varepsilon\}\\
&\leqslant P\{|X_n|\geqslant M\}+P\{|Y_n-Y|\geqslant\varepsilon/(2M)\}\\
&+P\{|Y|\geqslant M\}+P\{|X_n-X|\geqslant\varepsilon/(2M)\}<\tfrac{1}{4}\delta\times4\\
&=\delta,
\end{aligned}
$$

$\therefore \quad X_nY_n \xrightarrow{P} XY\ (n\to\infty)$.

27.设$X_n\xrightarrow{P}X$.而g是R^1上的连续函数，试证

$g(X_n)\xrightarrow{P}g(X)$.

证法一：对任给$\delta>0$，取$M>0$及N_1，使当$n>N_1$时有

$$P\{|X|\geqslant M\}<\tfrac{1}{8}\delta,\ P\{|X_n-X|\geqslant M\}<\tfrac{1}{8}\delta.$$

$g(x)$在$[-2M,2M]$上一致连续，则对任给$\varepsilon>0$,存在$\varepsilon_1>0$,使当$x_1,x_2\in[-2M,2M]$且$|x_1-x_2|<\varepsilon_1$时有

$$|g(x_1)-g(x_2)|<\varepsilon.$$

再取$N\geqslant N_1$，使当$n>N$时有

$$P\{|X_n-X|\geqslant\varepsilon_1\}<\frac{1}{3}\delta.$$

由于

$$\{|X|\geqslant 2M\}\cup\{|X_n|\geqslant 2M\}\subset$$
$$\subset\{|X|\geqslant M\}\cup\{|X|<M,|X_n|\geqslant 2M\}\subset$$
$$\subset\{|X|\geqslant M\}\cup\{|X_n-X|\geqslant M\},$$
$$\{|X|\leqslant 2M,\ |X_n|\leqslant 2M,\ |g(X_n)-g(X)|\geqslant\varepsilon\}$$
$$\subset\{|X_n-X|\geqslant\varepsilon_1\},$$

所以当$n>N$时有

$$P\{|g(X_n)-g(X)|\geqslant\varepsilon\}\leqslant$$
$$\leqslant P\{(|X|\geqslant 2M)\cup(|X_n|\geqslant 2M)\}+$$
$$+P\{|X|\leqslant 2M,\ |X_n|\leqslant 2M,\ |g(X_n)-g(X)|\geqslant\varepsilon\}$$
$$\leqslant P\{|X|\geqslant M\}+P\{|X_n-X|\geqslant M\}+P\{|X_n-X|\geqslant\varepsilon_1\}$$
$$\leqslant\frac{1}{3}\delta+\frac{1}{3}\delta+\frac{1}{3}\delta=\delta,$$

$$\therefore g(X_n)\xrightarrow{P}g(X),\ (n\to\infty).$$

证法二： $P(\Omega)=1$是有限测度，在实变函数论中曾得到，这时$X_n\xrightarrow{P}X$的充要条件是，对$\{X_n\}$的任一子序列$\{X_{n_k}\}$，都能找到其的一子序列$\{X_{n_{k_\nu}}\}$几乎处处收敛于X。(上题也可以用此定理证)对序列$\{g(X_n)\}$的任一子序列$\{g(X_{n_k})\}$。因为$X_n\xrightarrow{P}X$，由充要条件得，对$\{X_{n_k}\}$可找到其一子序列$\{X_{n_{k_\nu}}\}$，使$X_{n_{k_\nu}}\xrightarrow{a.e.}X$。由于g是$R^1$的连续函数，由此得$g(X_{n_{k_\nu}})\xrightarrow{a.e.}g(X)$。再由充要条件得$g(X_n)\xrightarrow{P}g(X)$。

28. 若$\{X_n\}$是单调下降的正随机变量序列，且$X_n\xrightarrow{P}0$，

证 明 $X_n \xrightarrow{a.s.} 0$。

证：由序列的单调下降性可得，当$n\to\infty$时X_n的极限存在，且

$$\{\lim_{n\to\infty} X_n<\varepsilon\}\supset\{X_k<\varepsilon\},$$

由 $X_k \xrightarrow{P} 0$ 得

$$1\geqslant P\{\lim_{n\to\infty} X_n<\varepsilon\}\geqslant \lim_{k\to\infty} P\{X_k<\varepsilon\}=1,$$

再由$X_n>0$及 ε 的任意性得

$$P\{\lim_{n\to\infty} X_n=0\}=1,\ \text{即} X_n\xrightarrow{a.s.}0.$$

29.若$X_1,X_2,\cdots$是独立随机变量序列，μ是整值随机变量，$P\{\mu=k\}=p_k$，且与$\{X_i\}$独立，求$\eta=X_1+\cdots+X_\mu$的特征函数。

解：设事件$B_1,B_2,\cdots,B_n$互不相容，$P(B_i)>0(i=1,2,\cdots,n)$，而且$\bigcup_{i=1}^{n} B_i=\Omega$，由全概率公式得

$$F(x)=P\{\xi<x\}=\sum_{i=1}^{n} P\{\xi<x|B_i\}P(B_i)$$

$$=\sum_{i=1}^{n} F(x|B_i)P(B_i).$$

所以有

$$E\xi=\int_{-\infty}^{\infty} x\mathrm{d}F(x)=\sum_{i=1}^{n} E(\xi|B_i)P(B_i).$$

此式称为全数学期望公式。由此并利用独立性得

$$f_\eta(t)=E(e^{it\eta})=\sum_{n=1}^{\infty}E[\exp(it\sum_{k=1}^{\mu}X_k)|\mu=n]P\{\mu=n\}$$

$$=\sum_{n=1}^{\infty}E\exp(it\sum_{k=1}^{n}X_k)p_n=\sum_{n=1}^{\infty}(\prod_{k=1}^{n}f_{X_k}(t))p_n.$$

30.若$f(t)$是非负定函数，试证：(1)$f(0)$是实的，且$f(0)\geqslant 0$；(2) $f(-t)=\overline{f(t)}$；(3) $|f(t)|\leqslant f(0)$。

证：因为$f(t)$是非负定的，故对任何实数$t_1,t_2,\cdots,t_n$，复数$\lambda_1,\lambda_2\cdots,\lambda_n$，恒有

$$\sum_{k=1}^{n}\sum_{j=1}^{n}f(t_k-t_j)\lambda_k\overline{\lambda_j}\geqslant 0.$$

(1)令$n=1$，$t_1=0$，$\lambda_1=1$。由非负定性条件得

$$\sum_{k=1}^{n}\sum_{j=1}^{n}f(t_k-t_j)\lambda_k\overline{\lambda_j}=f(0)\geqslant 0.$$

(2)令$n=2$，$t_1=0$，$t_2=t$得

$$0\leqslant\sum_{k=1}^{2}\sum_{j=1}^{2}f(t_k-t_j)\lambda_k\overline{\lambda_j}=$$

$$=f(0-0)\lambda_1\overline{\lambda_1}+f(0-t)\lambda_1\overline{\lambda_2}+f(t-0)\lambda_2\overline{\lambda_1}+f(t-t)\lambda_2\overline{\lambda_2}$$

$$=f(0)(|\lambda_1|^2+|\lambda_2|^2)+f(-t)\lambda_1\overline{\lambda_2}+f(t)\overline{\lambda_1}\lambda_2,$$

所以

$$f(-t)\lambda_1\overline{\lambda_2}+f(t)\overline{\lambda_1}\lambda_2$$

应该是实数。设

$$f(-t)=\alpha_1+i\beta_1,f(t)=\alpha_2+i\beta_2,\ \lambda_1\overline{\lambda_2}=\gamma+i\delta,$$

$\overline{\lambda_1}\lambda_2=\gamma-i\delta$,

代入上式并设虚部为0得

$$(\alpha_1-\alpha_2)\delta+(\beta_1+\beta_2)\gamma=0.$$

由γ，δ的任意性得

$$\alpha_1-\alpha_2=0,\ \beta_1+\beta_2=0,\ 即\ f(-t)=\overline{f(t)}.$$

(3)在(2)中令$\lambda_1=f(t)$，$\lambda_2=-|f(t)|$，得

$$2f(0)|f(t)|^2-|f(t)|^2|f(t)|-|f(t)|^2|f(t)|\geqslant 0,$$

若$|f(t)|\neq 0$，则得$f(0)\geqslant|f(t)|$；若$|f(t)|=0$，则由(1)中结果得$f(0)\geqslant|f(t)|$。

*31．证明：(1)若$g(x)$是连续函数，且$g(\mp\infty)=0$，又分布函数列$\{F_n(x)\}$弱收敛于分布函数$F(x)$，则

$$\int_{-\infty}^{\infty}g(x)dF_n(x)\to\int_{-\infty}^{\infty}g(x)dF(x).$$

(2)若分布函数$F(x)$对应的特征函数为$f(t)$，定义积分特征函数为

$$\hat{f}(t)=\int_0^t f(u)du=\int_{-\infty}^{\infty}\frac{e^{itx}-1}{ix}dF(x).$$

试证$F(x)$与$\hat{f}(t)$之间有一个一一对应。(3)若分布函数列$\{F_n(x)\}$弱收敛于分布函数$F(x)$，则相应的积分特征函数收敛：$\hat{f}_n(t)\to\hat{f}(t)$；反之，若$\hat{f}_n(t)$收敛到一个函数$\hat{h}(t)$，则存在一个分布函数使$F_n(t)\xrightarrow{W}F(x)$，且$\hat{f}(t)=\hat{h}(t)$。

证：本题参阅洛易甫著（梁文骐译）“概率论”上册第四章。该书中对分布函数作了推广定义：$F(x)$是x的非降函数，在R上处处左连续且$0\leqslant F(x)\leqslant 1$。如此定义的分布函数，当$F(-\infty)=0$，$F(+\infty)=1$时，则$F(x)$是一个随机变量（概率空间上的有限可测函数）的分布函数；当$F(-\infty)>0$

或$F(+\infty)<1$时，则$F(x)$仅是概率空间上一个可测函数的分布函数。本题中$F_n(x)\xrightarrow{W}F(x)$，弱收敛的极限函数$F(x)$是指以上推广意义下的分布函数。

(1)由$g(x)$连续且$\lim\limits_{x\to\pm\infty}g(x)=0$，可推得$|g(x)|\leqslant k<\infty$ $(x\in R^1)$，所以$\int_{-\infty}^{\infty}g(x)dF_n(x)$与$\int_{-\infty}^{\infty}g(x)dF(x)$积分存在且有限。我们有

$$\left|\int_{-\infty}^{\infty}g(x)dF_n(x)-\int_{-\infty}^{\infty}g(x)dF(x)\right|\leqslant$$

$$\leqslant\left|\int_{-\infty}^{\infty}g(x)dF_n(x)-\int_a^b g(x)dF_n(x)\right|+$$

$$+\left|\int_a^b g(x)dF_n(x)-\int_a^b g(x)dF(x)\right|$$

$$+\left|\int_{-\infty}^{\infty}g(x)dF(x)-\int_a^b g(x)dF(x)\right|=$$

$$=I_1+I_2+I_3. \tag{1}$$

I_1,I_2显然均不超过 $\mathrm{Sup}\ \{|g(x)|;\ x\overline{\in}[a,b]\}$。对任给$\varepsilon>0$，存在正数$M_1$，使当$x<-M_1$时有$|g(x)|<\frac{1}{3}\varepsilon$，又存在正数$M_2$，使当$x>M_2$时有$|g(x)|<\frac{1}{3}\varepsilon$。这时若取$a<-M_1$，$b>M_2$，则有

$$I_1<\frac{1}{3}\varepsilon,\ I_3<\frac{1}{3}\varepsilon. \tag{2}$$

在如上选取a,b时，使它们都是$F(x)$的连续点，则由海莱第二定理知，当n充分大时有

$$I_2<\frac{1}{3}\varepsilon. \tag{3}$$

由(1)—(3)得，当n充分大时有$I_1+I_2+I_3<\varepsilon$，此即

$$\lim_{n\to\infty}\int_{-\infty}^{\infty}g(x)dF_n(x)=\int_{-\infty}^{\infty}g(x)dF(x).$$

证明中不需要条件$F_n(-\infty)\to F(-\infty)$，$F_n(+\infty)\to F(+\infty)$，所以分布函数$F(x)$的全变差可以小于1。

(2) $\hat{f}(t)=\int_0^t f_F(u)\,du=\int_0^t du\int_{-\infty}^{\infty}e^{iux}dF(x)$。

由于$|e^{iux}|=1$对$P_F\times L[0,t]$测度可积，由富比尼定理知可交换积分次序，得

$$\hat{f}(t)=\int_{-\infty}^{\infty}\frac{e^{itx}-1}{ix}dF(x)。$$

首先特征函数$f(t)$与分布函数$F(x)$之间有一一对应关系，又对$\hat{f}(t)$求导得$f(t)$，对$f(t)$积分得$\hat{f}(t)$，所以$\hat{f}(t)$与$F(x)$之间也有一一对应关系。

(3) 设$F_n(x)\xrightarrow{W}F(x)$。由关系式$\lim_{x\to 0}\frac{e^{itx}-1}{ix}=t$知，$\frac{e^{itx}-1}{ix}$是$x\in R^1$的连续函数；又有$\lim_{x\to\pm\infty}\frac{e^{itx}-1}{ix}=0$，由(1)中结论得$\hat{f}_n(t)\to\hat{f}_F(t)$。

反之，设$\hat{f}_n(t)\to\hat{h}(t)$。由于$F_n(x)$是$\hat{f}_n(t)$所对应的分布函数，由海莱第一定理知，存在子序列$\{F_{n_k}(x)\}$，使$F_{n_k}(x)\xrightarrow{W}F(x)$。$F(x)$单调非降左连续且$0\leqslant F(x)\leqslant 1$，所以$F(x)$是推广意义下的分布函数，故由(1)中结论得$\hat{f}_{n_k}(t)\to\hat{f}_F(t)$且$\hat{f}_F(t)=\hat{h}(t)$。若$\{F_n(x)\}$有另一子列$\{F_{n_l}(x)\}$，使$F_{n_l}(x)\xrightarrow{W}F'(x)$，由(1)中结论得$\hat{f}_{n_l}(t)\to\hat{f}_{F'}(t)$且$\hat{f}_{F'}(t)=\hat{h}(t)$。所以$\hat{f}_F(t)=\hat{h}(t)=\hat{f}_{F'}(t)$。由(2)中结论知，分布函数被积分特征函数唯一

地确定，从而$F(x)=F'(x)$，

$$\therefore \quad F_n(x)\xrightarrow{W} F(x) \text{ 且 } \hat{f}(t)=\hat{h}(t)$$

32. 用特征函数法直接证明德莫哇佛—拉普拉斯积分极限定理。

证：即要证，若μ_n是 n 次贝努里试验中事件出现的次数，$0<p<1$，则对任意有限区间$[a,b]$，当$n\to\infty$时一致地有

$$P\left\{a\leqslant\frac{\mu_n-np}{\sqrt{npq}}<b\right\}\to\int_a^b\varphi(x)dx,$$

其中 $\varphi(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}$

因为μ_n服从二项分布$b(k;n,p)$，所以它的特征函数为$f_n(t)=(q+pe^{it})^n$，而$(\mu_n-np)/\sqrt{npq}$的特征函数为

$$g_n(t)=\left[q+p\exp\left(\frac{it}{\sqrt{npq}}\right)\right]^n\exp\left(-\frac{npit}{\sqrt{npq}}\right)$$

$$=\left[q\exp\left(-\frac{pit}{\sqrt{npq}}\right)+p\exp\left(\frac{qit}{\sqrt{npq}}\right)\right]^n$$

按台劳公式展开e^z

$$e^z=1+z+\frac{1}{2!}z^2+o(z^2),$$

则得

$$p\exp\left(\frac{qit}{\sqrt{npq}}\right)=p+it\sqrt{\frac{pq}{n}}-\frac{qt^2}{2n}+o\left(\frac{t^2}{n}\right)$$

$$q\exp\left(-\frac{pit}{\sqrt{npq}}\right)=q-it\sqrt{\frac{pq}{n}}-\frac{pt^2}{2n}+o\left(\frac{t^2}{n}\right).$$

代入$g_n(t)$得

$$g_n(t)=\left[1-\frac{t^2}{2n}+o\left(\frac{t^2}{n}\right)\right]^n\to e^{-\frac{1}{2}t^2}\quad(n\to\infty).$$

而 $e^{-\frac{1}{2}t^2}$ 是标准正态分布$N(0,1)$的特征函数，由逆极限定

理即可得要证的结论。

33. 若母体 ξ 的数学期望 $E\xi=m, D\xi=\sigma^2$，抽容量为 n 的子样求其平均值 $\overline{\xi}$，为使 $P\{|\overline{\xi}-m|<0.1\sigma\}\geqslant 95\%$，问 n 应取多大值？

解： 用林德贝格—勒维定理，记

$$\overline{\xi}-m=\frac{\sigma}{\sqrt{n}}\zeta_n,\quad 其中\zeta_n=\frac{1}{\sqrt{n}\,\sigma}\sum_{k=1}^{n}(\xi_k-m),\ 则$$

$$P\{|\overline{\xi}-m|<0.1\sigma\}=P\left\{\left|\frac{\sigma}{\sqrt{n}}\zeta_n\right|<0.1\sigma\right\}$$

$$=P\{|\zeta_n|<0.1\sqrt{n}\}\approx$$

$$\approx\frac{1}{\sqrt{2\pi}}\int_{-0.1\sqrt{n}}^{0.1\sqrt{n}}e^{-\frac{1}{2}t^2}dt$$

$$=\Phi(0.1\sqrt{n})-\Phi(-0.1\sqrt{n})=$$

$$=2\Phi(0.1\sqrt{n})-1=0.95,$$

$\Phi(0.1\sqrt{n})=0.975$，查表得 $0.1\sqrt{n}=1.96, n=385$。所以 n 至少应取385。

34. 若 $\{\xi_n, n=1,2,\cdots\}$ 为相互独立随机变量序列，具有相同分布 $P\{\xi_n=1\}=\frac{1}{2}, P\{\xi_n=0\}=\frac{1}{2}$，而 $\eta_n=\sum_{k=1}^{n}\frac{\xi_k}{2^k}$，试证 η_n 的分布收敛于[0,1]上的均匀分布。

证： ξ_k 的特征函数为

$$f_{\xi_k}(t)=\frac{1}{2}(e^{it}+1)=\cos\frac{t}{2}\,e^{\frac{1}{2}it}\ (k=1,2,\cdots),$$

所以 $\xi_k/2^k$ 的特征函数为

$$f_{\xi_k/2^k}(t)=\cos\frac{t}{2^{k+1}}\exp\left(\frac{it}{2^{k+1}}\right),$$

η_n 的特征函数为

$$f_{\eta_n}(t)=\cos\frac{t}{2^2}\cdot\cos\frac{t}{2^3}\cdots\cos\frac{t}{2^{n+1}}\times$$

$$\times\exp\left(it\left(\frac{1}{2^2}+\cdots+\frac{1}{2^{n+1}}\right)\right)$$

$$=\frac{1}{2^n}\cdot\frac{\sin\frac{t}{2}}{\sin\frac{t}{2^{n+1}}}\exp\left(\frac{it}{2}\left(1-\frac{1}{2^n}\right)\right)$$

$$\lim_{n\to\infty}f_{\eta_n}(t)=\frac{2}{t}\sin\frac{t}{2}\cdot e^{\frac{1}{2}it}=$$

$$=\frac{1}{it}(e^{it}-1)=f(t).$$

$f(t)$ 是$[0,1]$上均匀分布的特征函数，由逆极限定理得证。

35。用特征函数法证明二项分布的普阿松逼近定理。

证：二项分布的特征函数为

$$f_n(t)=(p_ne^{it}+q_n)^n=\left[1+\frac{np_n(e^{it}-1)}{n}\right]^n.$$

若当$n\to\infty$时$np_n\to\lambda$，则

$$np_n(e^{it}-1)\to\lambda(e^{it}-1)\ (n\to\infty).$$

所以

$$\lim_{n\to\infty}f_n(t)=\exp[\lambda(e^{it}-1)]=f(t).$$

$f(t)$是普阿松分布$p(\lambda)$的特征函数，由逆极限定理得证。

36。用特征函数法证明，普阿松分布当$\lambda\to\infty$时，渐近正态分布。

证：设ξ_λ服从参数为λ的普阿松分布，则

$$f_{\xi_\lambda}(t)=\exp[\lambda(e^{it}-1)].$$

令 $\eta_\lambda=\frac{\xi_\lambda-\lambda}{\sqrt{\lambda}}$，并在下式中按台劳公式展开$e^{iz}$得

$$f_{\eta_\lambda}(t)=e^{-i\sqrt{\lambda}t}\, f_{\xi_\lambda}\left(\frac{t}{\sqrt{\lambda}}\right)$$

$$=\exp[-i\sqrt{\lambda}t+\lambda(e^{it/\sqrt{\lambda}}-1)]$$

$$=\exp\left[-i\sqrt{\lambda}t+i\frac{t}{\sqrt{\lambda}}\lambda-\frac{t^2}{2}+O\left(\frac{1}{\sqrt{\lambda}}\right)\right]$$

$$=\exp\left[-\frac{t^2}{2}+O\left(\frac{1}{\sqrt{\lambda}}\right)\right]\to e^{-\frac{1}{2}t^2}\quad(\lambda\to\infty).$$

由逆极限定理得，普阿松分布当$\lambda\to\infty$时，渐近正态分布.

37. 若X_i是独立同分布随机变量序列，其分布分别为：(1)$[-a,a]$上均匀分布；(2)普阿松分布；(3)Γ-分布.记

$$Y_n=\sum_{i=1}^{n}(X_i-EX_i)\Big/\sqrt{\sum_{i=1}^{n}DX_i},$$

试计算Y_n的特征函数，并求$n\to\infty$时的极限.

解：(1)设X_i服从$[-a,a]$上的均匀分布，其特征函数为

$$f_{X_i}(t)=\frac{\sin ta}{ta}.\text{又 } EX_i=0, DX_i=\frac{1}{3}a^2, \sqrt{\sum_{i=1}^{n}DX_i}$$

$$=\sqrt{\frac{n}{3}a^2},\quad Y_n=\sum_{i=1}^{n}\frac{\sqrt{3}X_i}{\sqrt{n}\,a},\text{ 所以}$$

$$f_{Y_n}(t)=\left[f_{X_i}\left(\sqrt{\frac{3}{n}}\frac{t}{a}\right)\right]^n=\left(\frac{\sin\sqrt{3/n}\,t}{\sqrt{3/n}\,t}\right)^n$$

按台劳公式展开$\sin x$,

$$\sin x=x-\frac{x^3}{3!}+o(x^3),$$

得

$$f_{Y_n}(t)=\left[1-\frac{1}{6}\cdot\frac{3}{n}t^2+o\left(\frac{1}{n}\right)\right]^n$$

$$=\left[1-\frac{t^2}{2n}+o\left(\frac{1}{n}\right)\right]^n \to e^{-\frac{1}{2}t^2}\ (n\to\infty)$$

(2)设X_i服从参数为λ的普阿松分布，其特征函数为

$f_{X_i}(t)=\exp[\lambda(e^{it}-1)]$。又$EX_i=\lambda, DX_i=\lambda$，

$Y_n=\sum_{i=1}^{n}\frac{X_i-\lambda}{\sqrt{n\lambda}}$，所以

$$f_{Y_n}(t)=$$

$$=\left\{\exp\left(-i\frac{\sqrt{\lambda}}{\sqrt{n}}t\right)\exp\left[\lambda(e^{it/\sqrt{n\lambda}}-1)\right]\right\}^n$$

$$=\exp(-it\sqrt{\lambda n})\exp[n\lambda(e^{it/\sqrt{n\lambda}}-1)]$$

$$=\exp\left[-it\sqrt{\lambda n}+it\sqrt{n\lambda}-\frac{1}{2}t^2+O\left(\frac{1}{\sqrt{n}}\right)\right]$$

$$=\exp\left[-\frac{1}{2}t^2+O\left(\frac{1}{\sqrt{n}}\right)\right]\to e^{-\frac{1}{2}t^2}\ (n\to\infty).$$

(3)设X_i服从Γ-分布$G(\beta,\alpha)$，其特征函数为

$f_{X_i}(t)=\left(1-\frac{it}{\beta}\right)^{-\alpha}$。又$EX_i=\frac{\alpha}{\beta}$，$DX_i=\frac{\alpha}{\beta^2}$，

$Y_n=\sum_{i=1}^{n}\frac{X_i-\alpha/\beta}{\sqrt{n\alpha/\beta^2}}$，所以

$$f_{Y_n}(t)=\left[\exp\left(-i\sqrt{\frac{\alpha}{n}}\,t\right)\left(1-i\frac{t}{\sqrt{n\alpha}}\right)^{-\alpha}\right]^n$$

$$=\exp(-it\sqrt{n\alpha})\left(1-i\frac{t}{\sqrt{n\alpha}}\right)^{-n\alpha}.$$

两边取对数并按台劳公式展开得

$$\ln f_{Y_n}(t)=-it\sqrt{n\alpha}-n\alpha\ln\left(1-i\frac{t}{\sqrt{n\alpha}}\right)$$

$$= -it\sqrt{n\alpha} - n\alpha\left[-\frac{it}{\sqrt{n\alpha}} + \frac{t^2}{2n\alpha} + o\left(\frac{1}{n}\right)\right]$$

$$\to -\frac{1}{2}t^2\ (n\to\infty).$$

$$\therefore f_{Y_n}(t)\to e^{-\frac{1}{2}t^2}\ (n\to\infty).$$

38. 设X_n独立同分布，$P\{X_n=2^{k-2\ln k}\}=2^{-k}$ ($k=1,2,\cdots$)，则大数定律成立。

证： 由辛钦大数定律知，这时只要验证EX_i存在。

$$EX_i=\sum_{k=1}^{\infty}2^{-k}2^{k-2\ln k}=\sum_{k=1}^{\infty}4^{-\ln k}.$$

而

$$4^{-\ln k}=e^{-\ln 4\ln k}=(e^{\ln k})^{-\ln 4}=k^{-\ln 4},$$

又$\ln 4>1$，所以

$$EX_i=\sum_{k=1}^{\infty}k^{-\ln 4}<\infty,$$

从而大数定律成立。

39.若$\{X_i\}$是相互独立的随机变量序列，均服从$N(0,1)$，试证

$$W_n=\sqrt{n}\frac{X_1+\cdots+X_n}{X_1^2+\cdots+X_n^2}\ \text{及}\ U_n=\frac{X_1+\cdots+X_n}{\sqrt{X_1^2+\cdots+X_n^2}}$$

渐近正态分布$N(0,1)$。

证：(1)的证明。

$$W_n=\frac{\sqrt{n}\sum_{i=1}^{n}X_i}{\sum_{i=1}^{n}X_i^2}=\frac{\sum_{i=1}^{n}X_i\Big/\sqrt{n}}{\sum_{i=1}^{n}X_i^2\Big/n}=\frac{\xi_n}{\eta_n},$$

其中设$\xi_n=\sum_{i=1}^{n}X_i\Big/\sqrt{n}$，$\eta_n=\sum_{i=1}^{n}X_i^2\Big/n$。由$X_i\sim N(0,1)$可得$\xi_n\sim N(0,1)$。又$X_i$间独立，所以$X_i^2$间也独立，对$\{X_i^2\}$应用辛钦大数定律得$\eta_n=\frac{1}{n}\sum_{i=1}^{n}X_i^2\xrightarrow{P}EX_i^2=1$。由本章第25题(2)中结论知$W_n$渐近$N(0,1)$（$n\to\infty$）。

(2)的证明。

$$U_n=\frac{\sum_{i=1}^{n}X_i\Big/\sqrt{n}}{\sqrt{\sum_{i=1}^{n}X_i^2\Big/n}}=\frac{\xi_n}{\sqrt{\eta_n}},$$

ξ_n和η_n同(1)中设。由$\eta_n\xrightarrow{P}1$及本章第27题结论得$\sqrt{\eta_n}\xrightarrow{P}1$。与(1)同理得$U_n$渐近$N(0,1)$（$n\to\infty$）。

40。设$X_1,X_2,\cdots$是独立随机变量序列，均服从〔0,1〕均匀分布，令$Z_n=\left(\prod_{i=1}^{n}X_i\right)^{1/n}$，试证$Z_n\xrightarrow{P}c$，这里c是常数，并求c。

证： 取对数得

$$\ln Z_n=\frac{1}{n}\sum_{i=1}^{n}\ln X_i。$$

因为X_i独立同分布，所以$\ln X_i$也独立同分布。又

$$E\ln X_i=\int_0^1\ln x\cdot 1\cdot dx=(x\ln x-x)\Big|_0^1=-1。$$

由辛钦大数定律得

$\ln Z_n \xrightarrow{P} -1$，即有$Z_n \xrightarrow{P} e^{-1} = c\ (n\to\infty)$。

41．若$\{X_i\}$是独立同分布随机变量序列，$EX_i = m$，若$f(x)$是一个有界的连续函数，试证

$$\lim_{n\to\infty} E\left[f\left(\frac{X_1+\cdots+X_n}{n}\right)\right] = f(m).$$

证： 因为X_i独立同分布且$EX_i = m$，所以由于柯尔莫哥洛夫定理（独立同分布场合的强大数定律）得

$$\frac{X_1+\cdots+X_n}{n} \xrightarrow{a.s.} E(X_i) = m\ (n\to\infty).$$

又$|f(x)| \leqslant k, P(\Omega) = 1$有限，由控制收 定理和$f(x)$的连续性得

$$\lim_{n\to\infty} E\left[f\left(\frac{X_1+\cdots+X_n}{n}\right)\right] =$$

$$= E\left[\lim_{n\to\infty} f\left(\frac{X_1+\cdots+X_n}{n}\right)\right]$$

$$= E\left[f\left(\lim_{n\to\infty} \frac{X_1+\cdots+X_n}{n}\right)\right] = f(m).$$

42．若$\{X_i\}$是独立同分布、具有有限二阶矩的随机变量序列，试证

$$\frac{2}{n(n+1)}\sum_{i=1}^{n} iX_i \xrightarrow{P} EX_1.$$

证： 记$EX_1 = a, DX_1 = \sigma^2 < \infty$，则

$$E\left(\frac{2}{n(n+1)}\sum_{i=1}^{n} iX_i\right) = \frac{2}{n(n+1)}\sum_{i=1}^{n} iEX_i$$

$$=a\cdot\frac{2}{n(n+1)}\sum_{i=1}^{n}i=a,$$

$$D\left(\frac{2}{n(n+1)}\sum_{i=1}^{n}iX_i\right)=$$

$$=\frac{4}{n^2(n+1)^2}\sum_{i=1}^{n}i^2\sigma^2$$

$$=\sigma^2\cdot\frac{4}{n^2(n+1)^2}\cdot\frac{n(n+1)(2n+1)}{6}$$

$$=\frac{2(2n+1)\sigma^2}{3n(n+1)}\to 0\ (n\to\infty).$$

其中利用X_i间的独立性。由马尔可夫大数定律得

$$\frac{2}{n(n+1)}\sum_{i=1}^{n}iX_i\xrightarrow{P}a=EX_1.$$

43。设 $X_1,X_2,\cdots$相互独立，均服从柯西分布 $p(x)=\frac{1}{\pi}\frac{1}{1+x^2}$。试证它们不满足格涅坚科关于大数定律的充要条件（见本章习题9），即要指出，当$n\to\infty$时

$$E\frac{\left(\sum_{i=1}^{n}X_i\right)^2}{n^2+\left(\sum_{i=1}^{n}X_i\right)^2}\nrightarrow 0.$$

证：由第四章第49题知，当柯西分布的密度函数为$f(x)=\frac{1}{\pi}\frac{\lambda}{\lambda^2+(x-\mu)^2}$时，其特征函数为$e^{i\mu t-\lambda|t|}$。现诸 X_i密度函数为 $p(x)=\frac{1}{\pi}\frac{1}{1+x^2}$，则其特征函数为 $e^{-|t|}$，利用

独立性得$\sum_{i=1}^{n} X_i$的特征函数为$(e^{-|t|})^n = e^{-n|t|}$。由此得$\sum_{i=1}^{n} X_i$的密度函数为$\frac{1}{\pi} \cdot \frac{n}{n^2+x^2}$，从而

$$E\frac{\left(\sum_{i=1}^{n} X_i\right)^2}{n^2+\left(\sum_{i=1}^{n} X_i\right)^2}=$$

$$=\int_{-\infty}^{\infty}\frac{t^2}{n^2+t^2}\cdot\frac{1}{\pi}\frac{n}{n^2+t^2}dt$$

$$=\frac{1}{\pi}\int_{-\infty}^{\infty}\frac{nt^2}{(n^2+t^2)^2}dt\quad(t=nu)$$

$$=\frac{1}{\pi}\int_{-\infty}^{\infty}\frac{u^2}{(1+u^2)^2}du$$

$$=\frac{1}{\pi}\int_{-\infty}^{\infty}\left[\frac{1}{1+u^2}-\frac{1}{(1+u^2)^2}\right]du$$

$$=\frac{1}{\pi}\left[\operatorname{arctg}u-\frac{u}{2(1+u^2)}-\frac{1}{2}\operatorname{arctg}u\right]\Big|_{-\infty}^{\infty}$$

$$=\frac{1}{\pi}\left(\frac{\pi}{4}+\frac{\pi}{4}\right)=\frac{1}{2}\nrightarrow 0,\ (n\to\infty).$$

44.（魏尔斯特拉斯定理的概率论证明）设$f(x)$是$[0,1]$上连续函数，利用概率论方法证明，必存在多项式序列$\{B_n(x)\}$，在$[0,1]$上一致收敛于$f(x)$。

证：利用伯恩斯坦多项式

$$B_n(x)=\sum_{m=0}^{n}\binom{n}{m}x^m(1-x)^{n-m}f\left(\frac{m}{n}\right).$$

显然$B_n(0)=f(0)$，$B_n(1)=f(1)$，故只要考虑$(0,1)$中的x。任

取一贝努里试验$\widetilde{E}=E^{\infty}$，使事件A在每次试验E中出现的概率恰为x，x任意固定；并以η_n表前n次试验中A出现的总次数，则由全数学期望公式（见本章第29题解）得

$$
\begin{aligned}
f(x)-B_n(x)= \\
&=\sum_{m=0}^{n}\binom{n}{m}x^m(1-x)^{n-m}\left[f(x)-f\left(\frac{m}{n}\right)\right] \\
&=E\left[f(x)-f\left(\frac{\eta_n}{n}\right)\right] \\
&=P\left\{\left|\frac{\eta_n}{n}-x\right|<\delta\right\}\times \\
&\quad\times E\left\{\left[f(x)-f\left(\frac{\eta_n}{n}\right)\right]\middle|\left|\frac{\eta_n}{n}-x\right|<\delta\right\} \\
&\quad+P\left\{\left|\frac{\eta_n}{n}-x\right|\geqslant\delta\right\}\times \\
&\quad\times E\left\{\left[f(x)-f\left(\frac{\eta_n}{n}\right)\right]\middle|\left|\frac{\eta_n}{n}-x\right|\geqslant\delta\right\}. \quad (1)
\end{aligned}
$$

其中$\delta>0$为如下选定的数：由$f(x)$的连续性，对任意$\varepsilon>0$，存在$\delta>0$，使当$|x-y|<\delta$，$0\leqslant x,y\leqslant1$时有

$$|f(x)-f(y)|<\frac{1}{2}\varepsilon. \quad (2)$$

令 $M=\sup\limits_{1\geqslant x\geqslant0}|f(x)|$，由(1)，(2)得

$$|f(x)-B_n(x)|\leqslant1\cdot\frac{\varepsilon}{2}+P\left\{\left|\frac{\eta_n}{n}-x\right|\geqslant\delta\right\}\cdot2M. \quad (3)$$

由贝努里大数定律得$\lim\limits_{n\to\infty}P\left\{\left|\frac{\eta_n}{n}-x\right|\geqslant\delta\right\}=0$，从而得证

$$\lim_{n\to\infty}B_n(x)=f(x),\ (0\leqslant x\leqslant1).$$

为证上式中收敛的一致性，利用车贝晓夫不等式

$$P\left\{\left|\frac{\eta_n}{n}-x\right|\geqslant\delta\right\}\leqslant\frac{D\eta_n}{n^2\delta^2}=\frac{nx(1-x)}{n^2\delta^2}<\frac{1}{n\delta^2},$$

故当$n>4M/(\varepsilon\delta^2)$时，由上式及(3)立得

$$\sup_{1\geqslant x\geqslant 0}|f(x)-B_n(x)|<\varepsilon.$$

45. 设$\{X_i\}$是独立随机变量序列，试证$X_n\xrightarrow{a.e.}0$的充要条件为，对任意$\varepsilon>0$有$\sum_{n=1}^{\infty}P\{|X_n|\geqslant\varepsilon\}<\infty$。

证： 充分性。对任意 $\varepsilon>0$，记 $A_n=\{|X_n|\geqslant\varepsilon\}$，则题设变成$\sum_{n=1}^{\infty}P(A_n)<\infty$。由波雷尔—康特立引理(i)知有

$$P\left\{\bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty}(|X_k|\geqslant\varepsilon)\right\}=0, \tag{1}$$

而这正是$\{X_n(\omega)\}$以概率1收敛于$X(\omega)$的等价表示（见教本P281(15)式），所以$X_n\xrightarrow{a.e.}0$。

必要性。由波雷尔—康特立引理(ii)及$\{X_n\}$的独立性得，

$$\sum_{n=1}^{\infty}P\{|X_n|\geqslant\varepsilon\}=\infty \tag{2}$$

成立的充要条件是，

$$P\left\{\bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty}(|X_k|\geqslant\varepsilon)\right\}=1. \tag{3}$$

而$X_n\xrightarrow{a.e.}0$的等价表示为，对任意$\varepsilon>0$(1)式成立。(1)与(3)

是矛盾的，这说明若$X_n \xrightarrow{a.e.} 0$，则不能有(2)式成立，所以应有$\sum_{n=1}^{\infty} P\{|X_n| \geq \varepsilon\} < \infty$。

46. 试证独立同分布随机变量序列，若存在有限的四阶中心矩，则强大数定律成立。

证：由题设知，随机变量序列$\{X_n\}$独立同分布，$EX_n = a$，$E(X_n - a)^4 = c_4 < \infty$，所以$DX_n = \sigma^2 < \infty$。由马尔柯夫不等式得

$$P\left\{\left|\frac{1}{n}\sum_{i=1}^{n} X_i - a\right| \geq \varepsilon\right\} \leq \frac{1}{\varepsilon^4} E\left[\frac{1}{n}\sum_{i=1}^{n}(X_n - a)\right]^4$$

$$= \frac{1}{n^4 \varepsilon^4}\left\{E\left[\sum_{i=1}^{n}(x_i - a)^4\right] + \right.$$

$$\left. + C_4^2 E\left[\sum_{i \neq j}(x_i - a)^2 (x_j - a)^2\right]\right\}$$

$$= \frac{1}{n^4 \varepsilon^4}\left[nc_4 + 6 \cdot \frac{n(n-1)}{2}\sigma^4\right] \leq \frac{1 + 3\sigma^4}{n^2 \varepsilon^4}.$$

其中用到，由独立性得$E(x_i - a)(x_j - a)^3 = 0 (i \neq j)$；最后一步成立，是由于当$n$很大时有$c_4 < n$。因为$\sum_{n=1}^{\infty} \frac{1 + 3\sigma^4}{n^2 \varepsilon^4} < \infty$

$$\therefore \quad \sum_{n=1}^{n} P\left\{\left|\frac{1}{n}\sum_{i=1}^{n} X_i - a\right| \geq \varepsilon\right\} < \infty.$$

由此利用波雷尔—康特立引理可得

$$P\left\{\bigcap_{k=1}^{\infty}\bigcup_{n=k}^{\infty}\left|\frac{1}{n}\sum_{i=1}^{r} X_i - a\right| \geq \varepsilon\right\} = 0,$$

所以$\frac{1}{n}(X_1+\cdots+X_n)\xrightarrow{a.s.}0$，强大数定律成立。

由于$\{X_n\}$独立同分布，且由$c_4<\infty$可得$\sigma^2<\infty$，所以若改用用柯尔莫哥洛夫强大数定理可立得结论。

47．习题6的独立随机变量序列是否满足强大数定律？

解：(1)用反证法证其不满足强大数定律。$EX_k=0$，若强大数定律成立，则

$$\frac{1}{n}(X_1+\cdots+X_n)\xrightarrow{a.s.}0,\tag{1}$$

因而亦有

$$\frac{1}{n}\sum_{k=1}^{n-1}X_k=\frac{n-1}{n}\cdot\frac{1}{n-1}\sum_{i=1}^{n-1}X_i\xrightarrow{a.s.}0.\tag{2}$$

(1)－(2)得

$$\frac{1}{n}X_n\xrightarrow{a.s.}0.\tag{3}$$

但由X_n的取值知，总有$\left|\frac{X_n}{n}\right|=\frac{2^n}{n}>1$，这显然与(3)式矛盾，所以对$\{X_k\}$强大数定律不成立。

(2)$DX_k=1$，由于

$$\sum_{k=1}^{\infty}\frac{DX_k}{k^2}=\sum_{k=1}^{\infty}\frac{1}{k^2}<\infty,$$

所以$\{X_k\}$满足柯尔莫哥洛夫强大数定律。

(3)用反证法证$\{X_k\}$不满足强大数定律。$EX_k=0$，若$\frac{1}{n}(X_1+\cdots+X_n)\xrightarrow{a.s.}0$，则由几乎处处收敛的等价表示知，应同时有

$$p_k = P\left\{\bigcup_{n=k}^{\infty}\left(\left|\frac{1}{n}\sum_{i=1}^{n}X_i\right| \geqslant \frac{1}{2}\right)\right\}, \quad \lim_{n\to\infty} p_k = 0; \tag{4}$$

$$q_k = P\left\{\bigcap_{n=k}^{\infty}\left(\left|\frac{1}{n}\sum_{i=1}^{n}X_n\right| < \frac{1}{2}\right)\right\}, \quad \lim_{k\to\infty} q_k = 1. \tag{5}$$

为估计 p_k 的值，我们把其中并事件分解成两两不相容 事 件的并，为此令

$$A_k = \left\{\frac{1}{k}\sum_{i=1}^{k}X_i \geqslant \frac{1}{2}\right\},$$

$$A_n = \left\{\left|\frac{1}{t}\sum_{i=1}^{t}X_i\right| < \frac{1}{2}, t = k, k+1, \cdots, n-1;\right.$$

$$\left.\left|\frac{1}{n}\sum_{i=1}^{n}X_i\right| \geqslant \frac{1}{2}\right\}, \quad (n>k). \tag{6}$$

换句话说，$\omega \in A_n(n>k)$的充要条件是，不等式

$\left|\frac{1}{t}\sum_{i=1}^{t}X_i\right| < \frac{1}{2}(t \geqslant k)$中第一个不成立的是 $t = n$ 的那一个。

显然 $A_n A_m = \phi$。设 $\omega \in A_n$，则由(6)知必有 $X_n(\omega) \neq 0$，由X_n的定义知$X_n(\omega) = n$或$-n$，得

$$p_k = P\left\{\bigcup_{n=k}^{\infty}A_k\right\} = \sum_{n=k}^{\infty}p(A_n)$$

$$\geqslant \sum_{n=k+1}^{\infty}P\left\{\left(\left|\frac{1}{t}\sum_{i=1}^{t}X_t\right| < \frac{1}{2}, t = k, k+1\cdots,\right.\right.$$

$$\left.\left. n-1\right)\cap\left(X_n = n\text{或}-n\right)\right\}$$

$\{X_k\}$是独立随机变量序列，由此可证得 X_n 与$\frac{1}{t}\sum_{i=1}^{t}X_i$

$(t=k,\cdots,n-1)$也独立（方法同教本 P155定理证明），由此得

$$p_k \geq \sum_{n=k+1}^{\infty} P\left\{\bigcap_{t=k}^{n-1}\left|\frac{1}{t}\sum_{i=1}^{t}X_i\right|<\frac{1}{2}\right\} P\left\{X_n=n\text{或}-n\right\} \tag{7}$$

由(5)得，对于常数 $c(0<c<1)$，存在 k_0，使当 $k>k_0$时有

$q_k>c$，更有$P\left\{\bigcap_{t=k}^{n-1}\left(\left|\frac{1}{t}\sum_{i=1}^{t}X_i\right|<\frac{1}{2}\right)\right\}>c$。把它代入(7)

式，并利用X_n的定义得，当$k>k_0$时有

$$p_k>\sum_{n=k+1}^{\infty}c\cdot\frac{1}{\sqrt{n}}=\infty.$$

这与(4)式矛盾。所以$\{X_n\}$不服从强大数定律。

48．举例说明波雷尔—康特拉引理(i)之逆不成立。

解： 设$A_1\supset A_2\supset\cdots\supset A_n\supset\cdots$，且$P(A_n)=\frac{1}{n}$，则

$$P\left\{\overline{\lim_{n\to\infty}}A_n\right\}=P\left\{\bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty}A_k\right\}$$

$$=P\left\{\bigcap_{n=1}^{\infty}A_n\right\}=P\left\{\lim_{n\to\infty}A_n\right\}$$

$$=\lim_{n\to\infty}P(A_n)=\lim\frac{1}{n}=0.$$

引理(i)中另一结论是等价性结论，所以也成立。但条件不成立，事实上有

$$\sum_{n=1}^{\infty} P(A_n) = \sum_{n=1}^{\infty} \frac{1}{n} = \infty,$$

故引理(i)之逆不真。

49．设 $\{X_n\}$ 是相互独立且具有有限方差的随机变量序列，若

$$\sum_{n=1}^{\infty} \frac{DX_n}{n^2} < \infty, \tag{1}$$

则必有

$$\lim_{n\to\infty} \frac{1}{n^2} \sum_{k=1}^{n} DX_k = 0. \tag{2}$$

证：设(1)式成立，则对任给 $\varepsilon>0$，存在N，当 $n \geqslant N$ 时有

$$\sum_{k=n+1}^{\infty} \frac{DX_k}{k^2} < \varepsilon,$$

$$0 \leqslant \frac{1}{n^2} \sum_{k=1}^{n} DX_k = \frac{1}{n^2} \sum_{k=1}^{N} DX_k + \frac{1}{n^2} \sum_{k=N+1}^{n} DX_k$$

$$\leqslant \frac{1}{n^2} \sum_{k=1}^{N} DX_k + \sum_{k=N+1}^{\infty} \frac{DX_k}{k^2} \leqslant \frac{1}{n^2} \sum_{k=1}^{N} DX_k + \varepsilon.$$

对固定的N，当 n 充分大时，上式右端第一项可小于 ε，

$$\therefore \quad \lim_{n\to\infty} \frac{1}{n^2} DX_k = 0.$$

50．设 $f(x)$ 和 $g(x)$ 在闭区间[0,1]上连续，且满足 $0 \leqslant f(x) \leqslant cg(x)$，这里 c 是一个正常数，则成立

$$\lim_{n\to\infty}\int_0^1\cdots\int_0^1\frac{f(x_1)+\cdots+f(x_n)}{g(x_1)+\cdots+g(x_n)}dx_1\cdots dx_n=$$

$$=\frac{\int_0^1 f(x)dx}{\int_0^1 g(x)dx}.$$

证： 设独立随机变量序列 $\{\xi_n\}$ 均服从〔0,1〕均匀分布，由存在定理（见王梓坤著“随机过程” P6）知必存在概率空间(Ω,F,P)，使$\{\xi_n\}$是定义在同一概率空间(Ω,F,P)上的独立随机变量序列，且均服从〔0,1〕均匀分布。这时$\{f(\xi_n(\omega))\}$，$\{g(\xi_n(\omega))\}$分别也是定义在(Ω,F,P)上的独立同分布随机变量序列。记

$$\eta_n(\omega)=\frac{f(\xi_1(\omega))+\cdots+f(\xi_n(\omega))}{g(\xi_1(\omega))+\cdots+g(\xi_n(\omega))},$$

若有 ω 使 $(g(\xi_1(\omega))+\cdots+g(\xi_n(\omega))=0$，则由$0\leqslant f(x)\leqslant cg(x)$知，对此 ω 也有 $f(\xi_1\omega))+\cdots+f(\xi_n(\omega))=0$，对这样的$\omega$令 $\eta_n(\omega)=0$。由此得$0\leqslant\eta_n(\omega)\leqslant c$，且$\eta_n(\omega)$在$(\Omega,F,P)$上有限可测，从而它是随机变量。我们有

$$E\eta_n=\int_\Omega\frac{f(\xi_1(\omega))+\cdots+f(\xi_n(\omega))}{g(\xi_1(\omega))+\cdots+g(\xi_n(\omega))}P(d\omega)=$$

$$=\int_0^1\cdots\int_0^1\frac{f(x_1)+\cdots+f(x_n)}{g(x_1)+\cdots+g(x_n)}dx_1\cdots dx_n.$$

$$\lim_{n\to\infty}\int_0^1\cdots\int_0^1\frac{f(x_1)+\cdots+f(x_n)}{g(x_1)+\cdots+g(x_n)}dx_1\cdots dx_n$$

$$=\lim_{n\to\infty}\int_\Omega\eta_n(\omega)P(d\omega)$$

$$\overset{(1)}{=}\int_\Omega\lim_{n\to\infty}\frac{[f(\xi_1(\omega))+\cdots+f(\xi_n(\omega))]/n}{[g(\xi_1(\omega))+\cdots+g(\xi_n(\omega))]/n}\times P(d\omega)$$

$$\overset{(2)}{=} \frac{Ef(\xi_1)}{Eg(\xi_1)} = \frac{\int_0^1 f(x)dx}{\int_0^1 g(x)dx},$$

其中等式(1)成立是利用控制收敛定理；等式(2)成立是利用独立同分布的强大数定理，这是因为$f(x)$与$g(x)$有界，$Ef(\xi_1)$与$Eg(\xi_1)$均存在且有限。

51．直接证明柯尔莫哥洛夫不等式。

本题是要证，若$\xi_1,\xi_2,\cdots,\xi_n$是独立随机变量，方差有限，记$S_n^2=\sum_{k=1}^{n}D\xi_k$，则对任意$\varepsilon>0$，成立

$$P\left\{\max_{n\geq k\geq 1}\left|\sum_{j=1}^{k}(\xi_j-E\xi_j)\right|\geq\varepsilon\right\}\leq\frac{S_n^2}{\varepsilon^2}. \tag{1}$$

证： 若$S_n^2=0$，则$D\xi_j=0, P\{\xi_j=E\xi_j\}=1$，因而(1)式双方都化为0而(1)显然正确；故不妨设$S_n^2>0$。改写(1)为

$$P\left\{\max_{n\geq k\geq 1}\left|\sum_{j=1}^{k}(\xi_j-E\xi_j)\right|\geq\varepsilon S_n\right\}\leq\frac{1}{\varepsilon^2}. \tag{2}$$

简记$A=\left\{\max_{n\geq k\geq 1}\left|\sum_{j=1}^{k}(\xi_j-E\xi_j)\right|\geq\varepsilon S_n\right\}$，我们有

$$A=\bigcup_{k=1}^{n}\left\{\left|\sum_{j=1}^{k}(\xi_j-E\xi_j)\right|\geq\varepsilon S_n\right\}. \tag{3}$$

故(2)式表示：n个事件$\left\{\left|\sum_{j=1}^{k}(\xi_j-E\xi_j)\right|\geq\varepsilon S_n\right\}$，

$k=1,2,\cdots,n$ 中至少有一个出现的概率不大于$1/\varepsilon^2$.(3)式虽然已把事件A分解为n个事件的和，但此n个事件未必不相交，为了要得到一个不相交的分解，对整数 v, $1\leqslant v\leqslant n$,令

$$A_v=\left\{\left|\sum_{j=1}^{k}(\xi_j-E\xi_j)\right|<\varepsilon S_n, k=0,1,\cdots,v-1;\right.$$

$$\left.\left|\sum_{j=1}^{v}(\xi_j-E\xi_j)\right|\geqslant\varepsilon S_n\right\}. \quad (4)$$

换句话说，$\omega\in A_v$的充要条件是：n个不等式

$$\left|\sum_{j=1}^{k}(\xi_j-E\xi_j)\right|<\varepsilon S_n\ (k=1,2\cdots,n)$$

中，第一个不成立的是第 v 个.显然

$$A=\bigcup_{v=1}^{n}A_v;\quad A_vA_\mu=\phi\ (v\neq\mu).$$

故(2)式化为

$$\sum_{v=1}^{n}p(A_v)\leqslant\frac{1}{\varepsilon^2}. \quad (5)$$

欲证明(1)式，现只要在 $S_n^2>0$的假设下证明(5)成立就够了.定义 n 个随机变量

$$Y_v(\omega)=\begin{cases}1, & 若\omega\in A_v\\ 0, & 若\omega\notin A_v\end{cases}\quad(v=1,\cdots,n). \quad (6)$$

由于$A_v(v=1,2,\cdots,n)$互不相交，因而

$$\sum_{v=1}^{n}Y_v(\omega)\leqslant 1.$$

令 $W_k=\sum_{j=1}^{k}(\xi_j-E\xi_j)$，以 W_n^2 乘上式两边，并取数学期望，得

$$\sum_{v=1}^{n}E(Y_vW_n^2)\leqslant S_n^2 . \tag{7}$$

为了估计 $E(Y_vW_n^2)$，注意

$$\begin{aligned}E(Y_vW_n^2)= E(Y_vW_n^2) &+\\ +2E[Y_v(W_n-W_v)W_v]&+\\ +E[Y_v(W_n-W_v)^2]&,\end{aligned} \tag{8}$$

但 $W_n-W_v=\sum_{j=v+1}^{n}(\xi_j-E\xi_j)$ 只涉及 $\xi_{v+1},\cdots,\xi_n$；$Y_vW_v=Y_v\sum_{j=1}^{v}(\xi_j-E\xi_j)$ 只涉及 $\xi_1,\cdots,\xi_v$，由 $\xi_1,\cdots,\xi_n$ 独立，可以证明 W_n-W_v 与 Y_vW_v 也独立（方法同教本P155定理证明），由 $E(W_n-W_v)=0$ 得

$$\begin{aligned}&E[Y_v(W_n-W_v)W_v]\\ &\quad=E(Y_vW_v)E(W_n-W_v)=0.\end{aligned} \tag{9}$$

又由(6)，(4)及 W_v 的定义可得

$$Y_vW_v^2\geqslant Y_v\varepsilon^2S_n^2 . \tag{10}$$

实际上，对任意 ω，若 $\omega\notin A_v$，由(6)知(10)双方在此 ω 上均化为0，故此时(10)成立；若 $\omega\in A_v$，则由(6)知 $Y_v(\omega)=1$，再由(4)中最后一不等式即知此时(10)也成立。

综合(7)，(8)，(9)，(10)，并注意 $Y_v\geqslant0$ 即得

$$S_n^2 \geqslant \sum_{v=1}^{n} E(Y_v W_n^2) \geqslant \sum_{v=1}^{n} E(Y_v W_v^2)$$

$$\geqslant \sum_{v=1}^{n} E(Y_v \varepsilon^2 S_n^2) = \varepsilon^2 S_n^2 \sum_{v=1}^{n} EY_v$$

$$= \varepsilon^2 S_n^2 \sum_{v=1}^{n} P(A_v).$$

由假设 $S_n^2>0$，等式两端消去 S_n^2，即得(5)式成立。

52．对事件序列$\{A_n\}$，及整数序列 $1=n_1<n_2<\cdots$，证明

(1) $\bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} A_n = \bigcap_{k=1}^{\infty} \bigcup_{j=k}^{\infty} \bigcup_{n=n_j}^{n_{j+1}-1} A_n$；

(2)若 $\sum_{j=1}^{\infty} P\left\{ \bigcup_{n=n_j}^{n_{j+1}-1} A_n \right\} < \infty$，

则 $P\left\{ \overline{\lim_{n\to\infty}} A_n \right\} = 0$；

(3)对随机变量序列$\{\zeta_n\}$，若级数

$$\sum_{j=1}^{\infty} P\left\{ \max_{n_{j+1}>n\geqslant n_j} |\zeta_n| \geqslant \varepsilon \right\}$$

对一切$\varepsilon>0$收敛，则$P\{\zeta_n\to 0\}=1$。

证：(1)若$\omega \in \bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} A_n$，则 ω 就属于无穷多个A_n，即存在子序列$m_1<m_2<\cdots<m_i<\cdots$，使 $\omega\in A_{m_i}$，$i=1,2,\cdots$。对于每个m_i必存在j，使$n_j\leqslant m_i<n_{j+1}$，若记

$B_j=\bigcup\limits_{n=n_j}^{n_{j+1}-1}A_n$。则由此得$\omega\in B_j$。注意到$m_i$有无穷多个，从而得$\omega$必属于无穷多个$B_j$，

$$\therefore \omega\in\bigcap_{k=1}^{\infty}\bigcup_{j=k}^{\infty}B_j=\bigcap_{k=1}^{\infty}\bigcup_{j=k}^{\infty}\bigcup_{n=n_j}^{n_{j+1}-1}A_n.$$

反之，若$\omega\in$无穷多个B_j，则必有$\omega\in$无穷多个A_n，即

$$\bigcap_{k=1}^{\infty}\bigcup_{n=k}^{\infty}A_n\supset\bigcap_{k=1}^{\infty}\bigcup_{j=k}^{\infty}B_j.$$

由如上两步证明即得证(1)。

(2)．若$\sum\limits_{j=1}^{\infty}P\left\{\bigcup\limits_{n=n_j}^{n_{j+1}-1}A_n\right\}<\infty$，则由波雷尔—康特立引理得($B_j$意义同(1))$P(\overline{\lim\limits_{j\to\infty}}B_j)=0$，由(1)中结论

$\overline{\lim\limits_{n\to\infty}}A_n=\overline{\lim\limits_{n\to\infty}}B_n$可得，$P(\overline{\lim\limits_{n\to\infty}}A_n)=0$。

(3)对任给$\varepsilon>0$，记$A_n=\{|\zeta_n|\geqslant\varepsilon\}$，则

$$\{\max_{n_{j+1}>n\geqslant n_j}|\zeta_n|\geqslant\varepsilon\}=\bigcup_{n=n_j}^{n_{j+1}-1}A_n.$$

题设条件转化为$\sum\limits_{j=1}^{\infty}P\left\{\bigcup\limits_{n=n_j}^{n_{j+1}-1}A_n\right\}<\infty$，利用(2)中结论可得

$$P\left(\overline{\lim_{n\to\infty}}A_n\right)=P\left(\bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty}\{|\zeta_n|\geqslant\varepsilon\}\right)=0.$$

由此即得$\zeta_n\xrightarrow{a.s.}0$（参看本章第45题解）。

53. 若$\{\xi_k\}$是独立随机变量序列，方差有限，记

$$S_n=\sum_{k=1}^{n}(\xi_k-E\xi_k),\quad \eta_n=\frac{1}{n}S_n.$$

(1)利用柯尔莫哥洛夫不等式证明

$$p_m=P\left\{\max_{2^{m+1}>n\geq 2^m}|\eta_n|\geq\varepsilon\right\}\leq\frac{1}{(2^m\varepsilon)^2}\sum_{j<2^{m+1}}D\xi_j;$$

(2)对上述p_m，证明若$\sum_{k=1}^{\infty}\frac{D\xi_k}{k^2}<\infty$，则$\sum_{m=1}^{\infty}p_m$收敛；

(3)利用上题结果证明对$\{\xi_n\}$成立柯尔莫哥洛夫强大数定律。

证： (1)$p_m=P\{\max|S_n|>2^m\varepsilon,2^m\leq n<2^{m+1}\}$

$$\leq P\{\max_{2^{m+1}>n\geq 1}|S_n|\geq 2^m\varepsilon\}$$

$$\leq\frac{1}{(2^m\varepsilon)^2}\sum_{j=1}^{2^{m+1}-1}D\xi_j.$$

其中后一步由柯尔莫哥洛夫不等式得来。

$$(2)\ 0\leq\sum_{m=1}^{\infty}p_m\leq\frac{1}{\varepsilon^2}\sum_{m=1}^{\infty}\frac{1}{2^{2m}}\sum_{j=1}^{2^{m+1}-1}D\xi_j$$

$$=\frac{1}{\varepsilon^2}\sum_{j=1}^{\infty}D\xi_j\sum_{m=m(j)}^{\infty}\frac{1}{2^{2m}}.$$

其中后一步由交换求和次序得来，求和下限 $m(j)$ 表示满足$2^{m+1}>j$的最小正整数m.由$2^{m(j)+1}>j$得$2^{2m(j)+2}>j^2$，从而$1/2^{2m(j)}<4/j^2$.由此得

$$\sum_{m=1}^{\infty}p_m\leq\frac{1}{\varepsilon^2}\sum_{j=1}^{\infty}D\xi_j\sum_{k=0}^{\infty}\frac{1}{2^{2(m(j)+k)}}$$

$$=\frac{1}{\varepsilon^2}\sum_{j=1}^{\infty}\frac{D\xi_j}{2^{2m(j)}}\sum_{k=0}^{\infty}\frac{1}{2^{2k}}$$

$$\leqslant\frac{1}{\varepsilon^2}\sum_{j=1}^{\infty}D\xi_j\cdot\frac{4}{j^2}\cdot\frac{1}{1-1/4}$$

$$=\frac{16}{3\varepsilon^2}\sum_{j=1}^{\infty}\frac{D\xi_j}{j^2}.$$

由此可得，若$\sum_{k=1}^{\infty}\frac{D\xi_k}{k^2}<\infty$，则$\sum_{m=1}^{\infty}p_m$收 .

(3) 设$\sum_{k=1}^{\infty}\frac{D\xi_k}{k^2}<\infty$，则由(2)中结论可得

$$\sum_{m=0}^{\infty}P\left\{\max_{2^{m+1}>n\geqslant 2^m}|\eta_n|\geqslant\varepsilon\right\}=$$

$$=P\{|\eta_1|\geqslant\varepsilon\}+\sum_{m=1}^{\infty}p_m<\infty,$$

再由上题的结论即得 $P\{\eta_n\to 0\}=0$，即柯尔莫哥洛夫大数定律成立：

$$P\left\{\frac{1}{n}\sum_{k=1}^{n}(\xi_k-E\xi_k)\to 0\right\}=1.$$

54.(1)设$\{c_k\}$为常数列，令

$$s_n=\sum_{k=1}^{n}c_k, b_m=\sup\{|s_{m+k}-s_m|, k=1,2,\cdots\}$$

$$b=\inf\{b_m, m=1,2,\cdots\}.$$

试证$\sum_{k=1}^{\infty} c_k$收敛的充要条件是$b=0$；

(2)(Kronecker引理)对实数列$\{c_k\}$，若$\sum_{k=1}^{\infty}\frac{c_k}{k}$收敛，则$\frac{1}{n}\sum_{k=1}^{n} c_k \to 0$。

证： (1)设$\sum_{k=1}^{\infty} c_k$收敛，则由柯西判别准则知，对任意$\varepsilon>0$，存在正整数n_0，使$|c_{n_0+1}+\cdots+c_{n_0+k}|<\varepsilon$，此即$|s_{n_0+k}-s_{n_0}|<\varepsilon$，对一切$k=1,2,\cdots$成立，由此可得$0\leqslant b\leqslant b_{n_0}\leqslant\varepsilon$，由$\varepsilon$的任意性知$b=0$。反之，由$b=0$知对任意$\varepsilon>0$，存在正整数$n_1$，使$b_{n_1}<\varepsilon$，因而$|s_{n_1+k}-s_{n_1}|<\varepsilon$对一切$k=1,2,\cdots$成立，所以对任意$m>n>n_1$，

$$|s_m-s_n|\leqslant|s_m-s_{n_1}|+|s_n-s_{n_1}|<2\varepsilon,$$

仍由柯西判别准则知$\sum_{k=1}^{\infty} c_k$收敛。

(2)先证若$\{c_k\}$为常数列，$\lim_{k\to\infty} c_k=c$，则$\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n} c_k=c$。事实上，对$n>n_0$有

$$\left|\frac{1}{n}\sum_{k=1}^{n} c_k-c\right| = \left|\frac{1}{n}\sum_{k=1}^{n}(c_k-c)\right|\leqslant$$

$$\leqslant\left|\frac{1}{n}\sum_{k=1}^{n_0}(c_k-c)\right|+\left|\frac{1}{n}\sum_{k=n_0+1}^{n}(c_k-c)\right|$$

对任意$\varepsilon>0$，选n_0使当$k>n_0$时，有$|c_k-c|<\frac{1}{2}\varepsilon$，于是上式右方第二项对任意$n>n_0$总小于$\frac{n-n_0}{n}\cdot\frac{\varepsilon}{2}<\frac{\varepsilon}{2}$，右方第一项当$n$充分大后也小于$\frac{1}{2}\varepsilon$。

现证Kronecker引理。令

$$s_0=0,\quad s_n=\sum_{k=1}^{n}\frac{c_k}{k},\quad t_n=\sum_{k=1}^{n}c_k,\quad n=1,2,\cdots,$$

则 $c_k=k(s_k-s_{k-1})$，$k=1,2,\cdots$，故

$$t_{n+1}=\sum_{k=1}^{n+1}ks_k-\sum_{k=1}^{n+1}ks_{k-1}=-\sum_{k=1}^{n}s_k+(n+1)s_{n+1},$$

$$n=1,2,\cdots,$$

$$\frac{t_{n+1}}{n+1}=-\frac{n}{n+1}\cdot\frac{1}{n}\sum_{k=1}^{n}s_k+s_{n+1}.$$

因$\{s_n\}$收敛于有穷极限，由上段所证知$\left\{\frac{1}{n}\sum_{k=1}^{n}s_k\right\}$也收敛于同一极限，故

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}c_k=\lim_{n\to\infty}\frac{t_{n+1}}{n+1}$$

$$=-\lim_{n\to\infty}\frac{n}{n+1}\cdot\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}s_k+\lim_{n\to\infty}s_{n+1}=0.$$

*55．若$\{\xi_k\}$是独立随机变量序列，方差有限，记

$$S_n'=\sum_{k=1}^{n}\frac{\xi_k-E\xi_k}{k},$$

(1)利用柯尔莫哥洛夫不等式证明

$$P\left\{\max_{1\geqslant k\geqslant n}\left|S'_{m+k}-S'_m\right|\geqslant\varepsilon\right\}\leqslant\frac{1}{\varepsilon^2}\sum_{k=m+1}^{m+n}\frac{D\xi_k}{k^2};$$

(2)利用上题结论(1)证明：若$\sum\limits_{k=1}^{\infty}\frac{D\xi_k}{k^2}<\infty$，则 S'_n以概率 1 收敛；(3)利用 Kronecker 引理导出柯尔莫哥洛夫强大数定律。

证：定义 $\xi'_k=\xi_k-E\xi_k$，则$E\xi'_k=0$，$D\xi'_k=D\xi_k$。令

$$S'_n(\omega)=\sum_{k=1}^{n}\frac{1}{k}\xi'_k(\omega),$$

$$b_m(\omega)=\sup\left\{\left|S'_{m+k}(\omega)-S'_m(\omega)\right|,\ k=1,2,\cdots\right\},$$

$$b(\omega)=\inf\{b_m(\omega),\ m=1,2,\cdots\}.$$

易知$b_m(\omega)(m\geqslant1)$和$b(\omega)$都是随机变量。

(1)对任意$\varepsilon>0$及任二正整数n，m，由柯尔莫哥洛夫不等式得

$$P\left\{\max_{1\geqslant k\geqslant n}\left|S'_{m+k}(\omega)-S'_m(\omega)\right|\geqslant\varepsilon\right\}\leqslant$$

$$\leqslant\frac{1}{\varepsilon^2}\sum_{k=m+1}^{m+n}\frac{D\xi'_k}{k^2}=\frac{1}{\varepsilon^2}\sum_{k=m+1}^{m+n}\frac{D\xi_k}{k^2}.$$

(2)令$\eta_n=\max\limits_{1\geqslant k\geqslant n}\left|S'_{m+k}(\omega)-S'_m(\omega)\right|$，则

$\eta_n(\omega)\uparrow b_m(\omega)$，$\left\{b_m(\omega)>\varepsilon\right\}\subset\bigcup\limits_{n=1}^{\infty}\left\{\eta_n(\omega)\geqslant\varepsilon\right\}$，由P的下连续性得

$$P\{b_m(\omega)>\varepsilon\}\leqslant\frac{1}{\varepsilon^2}\sum_{k=m+1}^{\infty}\frac{D\xi_k}{k^2}.$$

由此及 $b_m(\omega)\geqslant b(\omega)$ 推知

$$P\{b(\omega)>\varepsilon\}\leqslant P\{b_m(\omega)>\varepsilon\}\leqslant\frac{1}{\varepsilon^2}\sum_{k=m+1}^{\infty}\frac{D\xi_k}{k^2}.$$

令 $m\to\infty$，由于 $\sum_{k=1}^{\infty}\frac{D\xi_k}{k^2}<\infty$，得 $P\{b(\omega)>\varepsilon\}=0$．从而

自 $\{b(\omega)>0\}=\bigcup_{n=1}^{\infty}\left\{b(\omega)>\frac{1}{n}\right\}$ 及P的下连续性得

$$P\{b(\omega)>0\}=\lim_{n\to\infty}P\left\{b(\omega)>\frac{1}{n}\right\}=0,$$

于是几乎处处 $b(\omega)=0$．由此利用上题(1)中结论得，S_n' 几乎处处收敛。

(3)由上题中 Kronecker 引理及上段(2)中结论得

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}\xi_k'=\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}(\xi_k-E\xi_k)=0\ (a.s.)$$

即柯尔莫哥洛夫强大数定律成立。

56．设 $X_1,X_2,\cdots$ 是独立随机变量序列，对它成立中心极限定理，则对 $\{X_n\}$ 成立大数定律的充要条件为

$D(X_1+\cdots+X_n)=o(n^2)$．

证： 充分性．设 $D(X_1+\cdots+X_n)=o(n^2)$，则由车贝晓夫不等式得

$$P\left\{\left|\frac{1}{n}\sum_{i=1}^{n}\left(X_i-EX_i\right)\right|\geqslant\varepsilon\right\}\leqslant\frac{D(X_1+\cdots+X_n)}{\varepsilon^2n^2}\to0$$

$(n\to\infty)$

所以大数定律成立。

必要性。设对 $\{X_n\}$ 成立中心极限定理，即对任意 $a>0$ 有

$$\lim_{n\to\infty} P\left\{\frac{1}{B_n}\left|\sum_{i=1}^{n}(X_i-EX_i)\right|<a\right\}=$$

$$=\frac{1}{\sqrt{2\pi}}\int_{-a}^{a} e^{-\frac{1}{2}t^2}dt; \tag{1}$$

又成立大数定律，即对任给 $\varepsilon>0$ 有

$$\lim_{n\to\infty} P\left\{\frac{1}{n}\left|\sum_{i=1}^{n}(X_i-EX_i)\right|<\varepsilon\right\}=1. \tag{2}$$

而 $$P\left\{\frac{1}{n}\left|\sum_{i=1}^{n}(X_i-EX_i)\right|<\varepsilon\right\}=$$

$$=P\left\{\frac{B_n}{n}\cdot\frac{1}{B_n}\left|\sum_{i=1}^{n}(X_i-EX_i)\right|<\varepsilon\right\}$$

$$=P\left\{\frac{1}{B_n}\left|\sum_{i=1}^{n}(X_i-EX_i)\right|<\varepsilon\cdot\frac{n}{B_n}\right\},$$

由此利用(1)，(2)两式可得，当 $n\to\infty$ 时应有 $\varepsilon n/B_n\to\infty$，即 $B_n/n\to 0$，从而 $D(X_1+\cdots+X_n)=o(n^2)$。

57. 设 $X_1,X_2,\cdots$ 是独立同分布随机变量序列，且 $\sum_{k=1}^{n}X_k/\sqrt{n}$ 对每一个 $n=1,2,\cdots$ 有相同分布，那么，若 $EX_1=0$，$DX_1=1$，则 X_1 必须是 $N(0,1)$ 变量。

证：若记 $f(t)$ 为 X_i 的特征函数，则由题设知，它同样

也是$\frac{i}{\sqrt{n}}\sum_{k=1}^{n}X_k$的特征函数，因当n=1时它化成$X_1$，所以

$$f(t)=E\exp\left\{\frac{i}{\sqrt{n}}\sum_{k=1}^{n}X_k t\right\}=E\left(\prod_{k=1}^{n}\exp\left\{i\frac{X_k}{\sqrt{n}}t\right\}\right)$$

$$=\left[f\left(\frac{t}{\sqrt{n}}\right)\right]^n. \quad (1)$$

此式对每个n均成立。把$f\left(\frac{t}{\sqrt{n}}\right)$展成幂级数，注意到

$f^{(k)}(0)=i^kEX_j^k$，而$EX_j=0, EX_j^2=DX_j-(EX_j)^2=1$，所以

$$f\left(\frac{t}{\sqrt{n}}\right)=1+\frac{1}{2}\left(\frac{it}{\sqrt{n}}\right)^2+o\left(\frac{1}{n}\right)$$

$$=1-\frac{t^2}{2n}+o\left(\frac{1}{n}\right). \quad (2)$$

(2)代入(1)得

$$\left[f\left(\frac{t}{\sqrt{n}}\right)\right]^n=\left[1-\frac{t^2}{2n}+o\left(\frac{1}{n}\right)\right]^n\to e^{-\frac{1}{2}t^2}$$

$$(n\to\infty).$$

由于对任意n均有$f(t)=[f(t/\sqrt{n})]^n$，所以$f(t)=e^{-\frac{1}{2}t^2}$，而$X_1\sim N(0,1)$。

58. 设$\{X_k\}$是独立随机变量序列，且X_k服从$N(0,2^{-k})$，试证序列$\{X_k\}$：(1)成立中心极限定理；(2)不满足费勒条件；(3)不满足林德贝格条件，从而说明林德贝格条件并不是中心极限定理成立的必要条件。

证：(1)由正态分布有再生性知，$\frac{1}{B_n}\sum_{k=1}^{n}X_k$ 对任意 n 服从N(0,1)，所以中心极限定理成立。

(2) $B_n^2 < \sum_{k=1}^{\infty} b_k^2 = \sum_{k=1}^{\infty}\frac{1}{2^k} = 1$，$B_n^2 \nrightarrow \infty (n\to\infty)$，

由费勒条件的等价条件知，$\{X_n\}$不满足费勒条件。

(3)由$a_k=0$，$B_n\uparrow 1$，取$\tau = 1$得

$$c = \lim_{n\to\infty}\frac{1}{B_n^2}\sum_{k=1}^{n}\int_{|x-a_k|>\tau B_n}(x-a_k)^2 dF_k(x)$$

$$\geqslant \int_{|x|>1} x^2 dF_1(x) > 0,$$

其中右端仅保留和式中第一项。由此知$\{X_n\}$不满足林德贝格条件。

59. 若$\{X_k\}$是独立随机变量序列，X_1服从$[-1,1]$均匀分布，对$k=2,3,\cdots$，X_k服从$N(0,2^{k-1})$，证明对$\{X_k\}$成立中心极限定理，但不满足费勒条件。

证：$B_n^2 = \frac{1}{3} + \sum_{k=2}^{n} 2^{k-1} = \frac{1}{3} + \frac{2(1-2^{n-1})}{1-2}$

$$= 2^n - 1\frac{2}{3} \to \infty,$$

但
$$\frac{b_n^2}{B_n^2} = \frac{2^{n-1}}{2^n - 1\frac{2}{3}} \to \frac{1}{2} \neq 0 \ (n\to\infty),$$

所以费勒条件不满足。$\sum_{k=1}^{n}X_k$的特征函数为

$$f(t)=\prod_{k=1}^{n} f_{X_k}(t)=\frac{\sin t}{t}\exp\left\{-\frac{1}{2}(2^n-2)t^2\right\},$$

由此得 $\eta_n=\frac{1}{\sqrt{2^n-1\frac{2}{3}}}\sum_{k=1}^{n}X_k$的特征函数为

$$g_n(t)=\sin\frac{t}{\sqrt{2^n-1\frac{2}{3}}}\cdot\frac{\sqrt{2^n-1\frac{2}{3}}}{t}\times$$

$$\times\exp\left\{-\frac{1}{2}(2^n-2)t^2/(2^n-1\frac{2}{3})\right\}$$

$$\to e^{-\frac{1}{2}t^2}\quad(n\to\infty),$$

由逆极限定理知$\eta_n\xrightarrow{L}N(0,1)$，所以中心极限定理成立。

60．在普阿松试验中，第 i 次试验时事件 A出现的概率为P_i，不出现的概率为q_i，各次试验是独立的，以 ν_n 记前n次试验中事件A 出现的次数，试证：(1)$(\nu_n-E\nu_n)/n\xrightarrow{P}0$；(2) 对 $\left(\nu_n-\sum_{i=1}^{n}p_i\right)\Big/\sqrt{\sum_{i=1}^{n}p_iq_i}$ 成立中心极限定理的充要条件是$\sum_{i=1}^{\infty}p_iq_i=+\infty$。

证：(1)设$0\leqslant x\leqslant1$，则$y=x(1-x)=x-x^2$在$x=\frac{1}{2}$处取得极大值$\frac{1}{4}$，置$x=p_i$得$p_iq_i\leqslant\frac{1}{4}$.利用独立性得

$D\nu_n=\sum_{i=1}^{n}p_iq_i\leqslant\frac{1}{4}n$，再由车贝晓夫不等式得

$$P\left\{\left|\frac{\nu_n-E\nu_n}{n}\right|\geqslant\varepsilon\right\}\leqslant\frac{1}{\varepsilon^2}\frac{D\nu_n}{n^2}\leqslant\frac{1}{4\varepsilon^2n}\to0,$$

$$\therefore \qquad (v_n - Ev_n)/n \xrightarrow{P} 0 \ (n\to\infty).$$

(2)记 $v_n = X_1 + X_2 + \cdots + X_n$，这里

$$X_i = \begin{cases} 1, & 第\ i\ 次成功 \\ 0, & 第\ i\ 次失败, \end{cases} \qquad P\{X_i = 1\} = p_i,$$

$$P\{X_i = 0\} = q_i.$$

充分性.设$B_n = Dv_n = \sum_{i=1}^{n} p_i q_i \to \infty$，由于 X_i 仅取 0，1 两值，则 $P\{|X_i - p_i| \geqslant 2\} = 0$。由于 $B_n \to \infty$，故对任意 $\tau > 0$ 存在N，使当 $n > N$ 时 $B_n \geqslant 2/\tau$，即 $\tau B_n \geqslant 2$，所以 $P\{|X_i - p_i| \geqslant \tau B_n\} = 0$，从而当 $n > N$ 时

$$\frac{1}{B_n^2} \sum_{i=1}^{n} \int_{|x_i - p_i| \geqslant \tau B_n} (x - p_i)^2 dF_i(x)$$

$$= \frac{1}{B_n^2} \sum_{i=1}^{N} \int_{|x_i - p_i| \geqslant \tau B_n} (x - p_i)^2 dF_i(x)$$

$$\leqslant \frac{1}{B_n^2} \sum_{i=1}^{N} p_i q_i \leqslant \frac{N}{4B_n^2} \to 0 \ (n\to\infty),$$

即林德贝格条件成立。所以对 $\left(v_n - \sum_{i=1}^{n} p_i\right) \Big/ \sqrt{\sum_{i=1}^{n} p_i q_i}$ 中心极限定理成立。

必要性.设$\sum_{i=1}^{\infty} p_i q_i < \infty$，记$r_i = \min(p_i, q_i)$，则 $p_i q_i = r_i(1 - r_i) \geqslant \frac{1}{2} r_i$，所以

$$\sum_{i=1}^{\infty} r_i \leqslant \sum_{i=1}^{\infty} 2p_i q_i < \infty.$$

由此得 $\exp\left\{-\sum_{i=1}^{\infty} r_i\right\} > 0$，即 $\prod_{i=1}^{\infty} \exp\{-r_i\} > 0$。又 $\exp\{-r_i\} < 1 - r_i$，所以

$$c = \prod_{i=1}^{\infty}(1-r_i) > \prod_{i=1}^{\infty}\exp\{-r_i\} > 0.$$

以 $i_1 < i_2 < \cdots$ 记一切满足条件 $p_i \geqslant \frac{1}{2}$ 的 i 所构成的子序列，记 $Z_n = \left(v_n - \sum_{i=1}^{n} p_i\right) \Big/ \sqrt{\sum_{i=1}^{n} p_i p_i}$，以 m_n 记 $i_1, i_2, \cdots$ 中不超过 n 的个数，则有

$P\{v_n = m_n\} \geqslant P\{X_j = 1, j = i_1, \cdots, i_{m_n};\ X_j = 0,$ 对其它的 $j \leqslant n\} = \prod_{i=1}^{n}(1-r_i) \geqslant c > 0$。

于是若记 $d_n = \left(m_n - \sum_{i=1}^{n} p_i\right) \Big/ \sqrt{\sum_{i=1}^{n} p_i q_i}$，则对一切 n 有 $P\{Z_n = d_n\} \geqslant c$ 成立。由于 $c > 0$ 与 n 无关，所以 Z_n 不能依分布收敛到 $N(0,1)$，中心极限定理不成立。

61．设 $\{X_k\}$ 独立，X_k 服从 $[-k, k]$ 均匀分布，问对 $\{X_k\}$ 能否用中心极限定理？

解： 中心极限定理成立，因为这时可以直接证得林德贝格条件成立。

$a_k = EX_k = 0$，$DX_k = \frac{1}{3}k^2$，

$$B_n^2 = \sum_{k=1}^{n} DX_k = \frac{1}{3}\sum_{k=1}^{n} k^2 = \frac{1}{18}n(n+1)(2n+1).$$

对任意$\tau>0$，有

$$\frac{1}{B_n^2}\sum_{k=1}^{n}\int_{|x|>\tau B_n}(x-a_k)^2dF_k(x)\leqslant$$

$$\leqslant\frac{1}{B_n^2\cdot\tau B_n}\sum_{k=1}^{n}\int_{|x|>\tau B_n}|x|x^2f_k(x)dx$$

$$\leqslant\frac{1}{\tau B_n^3}\sum_{k=1}^{n}\int_{-k}^{k}|x|^3\cdot\frac{1}{2k}dx=\frac{1}{4\tau B_n^3}\sum_{k=1}^{n}k^3$$

$$=\frac{1}{4\tau}\cdot\frac{1}{4}n^2(n+1)^2\Big/\left[\frac{1}{18}n(n+1)(2n+1)\right]^{3/2}$$

$$=\frac{18^{3/2}}{16\tau}\cdot\frac{1}{\sqrt{n}}\left(1+\frac{1}{n}\right)^2\Big/\left[\left(1+\frac{1}{n}\right)\left(2+\frac{1}{n}\right)\right]^{3/2}\to 0$$

$$(n\to\infty)$$

所以林德贝格条件成立。

62．试问对下列独立随机变量序列，李雅普诺夫定理是否成立？(1) X_k: $\begin{pmatrix}-\sqrt{k} & \sqrt{k}\\ \frac{1}{2} & \frac{1}{2}\end{pmatrix}$；

(2) X_k: $\begin{pmatrix}-k^{\alpha} & 0 & k^{\alpha}\\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3}\end{pmatrix}$，$\alpha>0$。

解：(1) $EX_k=0$，$DX_k=\frac{1}{2}(-\sqrt{k})^2+\frac{1}{2}(\sqrt{k})^2$ $=k$，$B_n^2=1+2+\cdots+n=\frac{1}{2}n(n+1)$。取$\delta=1$，则

$$\frac{1}{B_n^{2+1}}\sum_{k=1}^{n}E|\xi_k-0|^{2+1}=$$

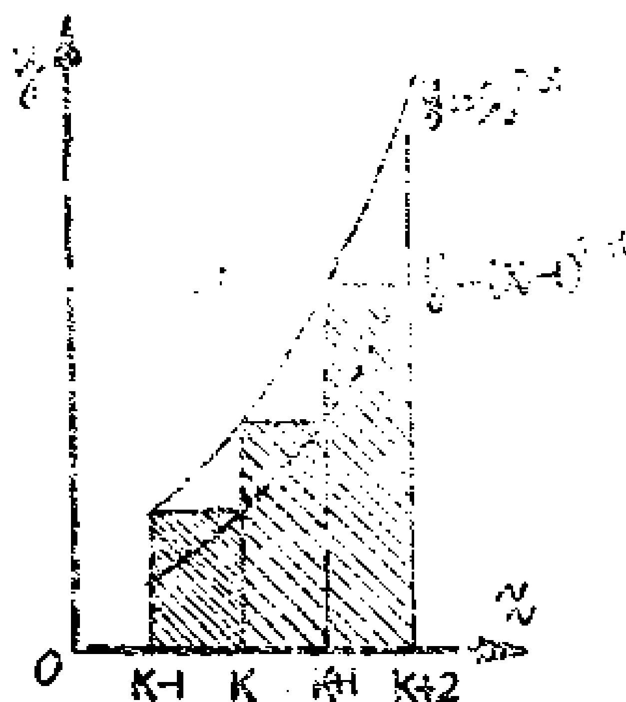


图 16

$$=\frac{1}{B_n^3}\sum_{k=1}^{n} k^{3/2} \leqslant \frac{n \cdot n^{3/2}}{\left[\frac{1}{2}n(n+1)\right]^{3/2}}$$

$$=\frac{2\sqrt{2}}{\sqrt{n}}\left(\frac{n}{n+1}\right)^{3/2} \to 0 \ (n\to\infty).$$

所以李雅普诺夫定理成立。

(2) $EX_k=0$, $DX_k=\frac{2}{3}k^{2\alpha}$, 可得(参看图16)

$$B_n^2=\frac{2}{3}\sum_{k=1}^{n} k^{2\alpha} \geqslant \frac{2}{3}\int_1^n (x-1)^{2\alpha}dx$$

$$=\frac{2}{3(2\alpha+1)}(x-1)^{2\alpha+1}\Big|_1^n$$

$$=\frac{2}{3(2\alpha+1)}(n-1)^{2\alpha+1}=c(n-1)^{2\alpha+1},$$

由$\alpha>0$得常数$c>0$, 取$\delta=1$得(参看图16)

$$\frac{1}{B_n^{2+1}}\sum_{k=1}^{n}E|X_k-0|^{2+1}=$$

$$=\frac{1}{B_n^3}\sum_{k=1}^{n}\frac{2}{3}k^{3\alpha}\leqslant\frac{2}{3B_n^3}\int_1^n x^{3\alpha}dx$$

$$\leqslant\frac{2}{3\sqrt{c^3}}\cdot\frac{1}{\sqrt{(n-1)^{3(2\alpha+1)}}}\times$$

$$\times\left(\frac{1}{3\alpha+1}n^{3\alpha+1}-\frac{1}{3\alpha+1}\right)$$

$$\leqslant\frac{2}{3(3\alpha+1)\sqrt{c^3}}\cdot\frac{1}{\left(1-\frac{1}{n}\right)^{3\alpha+1}}\cdot\frac{1}{\sqrt{n-1}}\to 0$$

$(n\to\infty)$。

所以李雅普诺夫定理成立。

63。求证：当$n\to\infty$时，

$$\frac{(\frac{1}{2}n)^{\frac{1}{2}n}}{\Gamma(\frac{1}{2}n)}\int_0^{1+t\sqrt{2/n}}y^{\frac{1}{2}n-1}e^{-\frac{1}{2}ny}dy\to$$

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{t}e^{-\frac{1}{2}y^2}dy.$$

证：设独立随机变量序列$\{\xi_n\}$都服从Γ-分布$G(\frac{1}{2},\frac{1}{2})$，$E\xi_1=1,D\xi_1=2$。由第四章第51题知$\Gamma$-分布具有再生性，

得$\eta_n=\sum_{k=1}^{n}\xi_k$服从$\Gamma$-分布$G\left(\frac{1}{2},\frac{n}{2}\right)$，且$E\eta_n=n,D\eta_n=2n$。

由于序列$\{\xi_n\}$独立同分布且方差有限，由林德贝格-勒维定理知，中心极限定理成立，据此把Γ-分布$G\left(\frac{1}{2},\frac{n}{2}\right)$的密

度函数代入计算得

$$P\{(\eta_n - n)/\sqrt{2n} < t\} = p\{\eta_n < n + t\sqrt{2n}\}$$

$$= \int_0^{n+t\sqrt{2n}} \frac{(\frac{1}{2})^{\frac{1}{2}n}}{\Gamma(\frac{1}{2}n)} x^{\frac{1}{2}n-1} e^{-\frac{1}{2}x} dx$$

(令$x = ny$)

$$= \int_0^{1+t\sqrt{2/n}} \frac{(\frac{1}{2}n)^{\frac{1}{2}n}}{\Gamma(\frac{1}{2}n)} y^{\frac{1}{2}n-1} e^{-\frac{1}{2}ny^2} dy$$

$$\to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-\frac{1}{2}y} dy \quad (n \to \infty).$$

64. 设$\lambda > 0$，独立随机变量序列$\{X_k\}$，对一切k，X_k以概率$\frac{1}{2}$分别取值$\pm k^{\lambda}$，试证：当$\lambda < \frac{1}{2}$时大数定律成立，对$\lambda \geqslant 1$则不成立。

证：当$\lambda < \frac{1}{2}$时见本章第 4 题。

当$\lambda \geqslant 1$时，若序列$\{X_k\}$服从大数定律，则应有

$$\frac{1}{n}(X_1 + \cdots + X_n) \xrightarrow{P} 0.$$

从而

$$\frac{X_n}{n} = \frac{1}{n}\sum_{k=1}^{n} X_k - \frac{n-1}{n} \cdot \frac{1}{n-1}\sum_{k=1}^{n-1} X_k \xrightarrow{P} 0$$

但由X_n的取值知总有

$$\left|\frac{X_n}{n}\right| = \left|\frac{n^{\lambda}}{n}\right| \geqslant \left|\frac{n}{n}\right| = 1,$$

这显然矛盾。故当$\lambda \geqslant 1$时大数定律不成立。

65. 求证：当$n\to\infty$时，$e^{-n}\sum_{k=0}^{n}\frac{n^k}{k!}\to\frac{1}{2}$。

证：设独立随机变量序列$\{\xi_n\}$有相同的普阿松分布，且参数为$\lambda=1$，由于普阿松分布具有再生性，所以$\eta_n=\sum_{k=1}^{n}\xi_k$是服从参数$\lambda=n$的普阿松分布，

$$P\{\eta_n=i\}=\frac{n^i}{i!}e^{-n}(i=0,1,2,\cdots),E\eta_n=D\eta_n=n.$$

由于$\{\xi_n\}$独立同分布且方差有限，由林德贝格—勒维定理知中心极限定理成立，所以

$$P\{(\eta_n-n)/\sqrt{n}\leqslant 0\}=P\{\eta_n\leqslant n\}$$

$$=e^{-n}\sum_{k=0}^{n}\frac{n^k}{k!}\to\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{0}e^{-\frac{1}{2}t^2}dt=\frac{1}{2}.$$

目　录

第二册　数理统计

第一章　引论 …… 1
§1. 统计学 …… 1
§2. 数理统计的基本概念 …… 3
§3. 求估计量的两种常用方法 …… 10
习题 …… 20
第二章　抽样分布 …… 22
§1. 正态母体子样的线性函数的分布 …… 23
§2. χ^2-分布 …… 25
§3. t-分布和 F-分布 …… 31
§4. 正态母体子样均值和方差的分布 …… 38
习题 …… 41
第三章　假设检验 (I) …… 43
§1. 引言 …… 43
§2. 正态母体参数的检验 …… 49
§3. 正态母体参数的置信区间 …… 65
§4. 多项分布的 χ^2-检验 …… 68
§5. 广义似然比检验 …… 82
习题 …… 86
第四章　线性统计推断 …… 89
§1. 最小二乘法 …… 89
§2. 回归分析 …… 113
§3. 方差分析 …… 123
习题 …… 140

第一章　引　论

§1. 统　计　学

统计学是一门关于数据资料的收集、整理、分析和推断的科学. 但人们常常将统计这一概念误解为大量数据资料的收集以及对这些数据作一些简单的运算(如求和、求平均值、求百分比等)或用图表、表格等形式把它们表示出来. 其实这些工作仅是统计学工作的非主要部分. 统计学还包括怎样设计试验、采集数据以及怎样对获得的数据进行分析、推断等其它许多工作.

随着研究随机现象规律性的科学——概率论的发展，应用概率论的结果更深入地分析研究统计资料，通过对某些现象的频率的观察来发现该现象的内在规律性，并作出一定精确程度的判断和预测；将这些研究的某些结果加以归纳整理，逐渐形成一定的数学概型. 这些组成了数理统计的内容.

数理统计的方法及考虑的问题不同于一般的资料统计，它更侧重于应用随机现象本身的规律性来考虑资料的收集、整理和分析，从而找出相应的随机变量的分布律或它的数字特征. 由于大量的随机试验必能呈现出它的规律性，因而从理论上讲，只要对随机现象进行足够多次观察，被研究的随机现象的规律性一定能清楚地呈现出来，但是实际上所允许的观察永远只能是有限的，有时甚至是少量的. 因此我们所关心的问题是怎样有效地利用有限的资料，便能去掉那些由于资料不足所引起的随机干扰，而把那些实质性的东西找出来. 一个好的统计方法就在于能有效地利用所获得的资料，尽可能作出精确而可靠的结论.

在数理统计里，不是对所研究的全部对象进行观察，而是抽取其中的部分进行观察、获得数据(即采样)，并通过这些数据来对所研究的全体进行推断. 由于推断是基于采样数据，而采样数据又不能包含研究对象的全部信息，因此由此所获得的结论必然会包含不定性. 概率是这种不定性的度量. 造成不定性的原因可分为两类：(1) 由于采样数据的随机性所引起的不定性；(2) 由于我们对系统真实状态的"无知"造成的不定性. 数理统计工作者的任务就是要分辨这两种不定性. 下面举一例来说明.

某元件厂生产了一批三极管，共一百万只，每十只装成一盒，共得十万盒. 现有某仪器厂需向该元件厂购买此三极管一百盒，准备安装在某种仪表上. 每台仪表需用此三极管十只，恰好是一台仪表需一盒三极管，一百盒可供装一百台. 但是该仪表对三极管有一定的质量要求，要求十只中至少有八只是一级品，其余的可以是二级品，否则仪表将不能稳定工作. 此时仪器厂对这批三极管就面临两种不定性需要分辨：(1) 元件厂生产的十万盒三极管，对仪器厂来说是满意的(即一盒中至少有八只是一级品)盒子所占比例 p 是多少？(2) 由于有十万盒三极管，现在仅购买其中的一百盒，因而就面临着另一种不定性. 即假使已知此十万盒中，满意的盒子所占的比例为 p，又怎样能确定买来的一百盒中，满意的占多少比例呢？例如 $p=0.99$，即十万盒中大约有九万九千盒是满意的，这个比例对仪器厂来说应该是很好了，但也有可能发生这样的意外，即仪器厂所购买的一百盒全部落在不满意的大约一千盒之中.

第一种不定性是不知道 p，是我们对系统真实状态的"无知"；而后一种不定性是由于所谓"随机性"造成的. 为了改善这不定性，仪器厂可要求元件厂对这批三极管的质量进行测试，也就是要求抽取部分三极管进行测试，通过这部分中一级品所占的比例(额

率)来对 p 的真实值进行推断. 当然我们不能完全精确地决定 p, 但是我们可以希望获得一个(在某种意义下)比较好的判断. 这就涉及到怎样设计试验,决定观察的数目,和怎样利用试验观察的结果作出一个"好的"推断等, 这些都是数理统计所要研究的问题. 至于在已知 p 的条件下, 第二种不定性的程度已在概率论基础部分作过讨论.

数理统计研究的内容随着科学技术和生产的不断发展而逐步扩大,但概括地说可以分为两大类: (1) 试验的设计和研究, 即研究如何更合理更有效地获得观察资料的方法; (2) 统计推断,即研究如何利用一定的资料对所关心的问题作出尽可能精确、可靠的结论. 当然这两部分有密切联系, 在实际应用中更应前后兼顾. 但因限于篇幅,本书将只讨论统计推断.

数理统计方法提供了在不定性占优势的情况下进行判断的工具,它为许多科学研究部门和生产、经济部门的工作者所采用. 它不仅为提高产量、质量起直接的推动作用,而且也提供了从大量现象中发现某些事物发展规律的方法.

§2. 数理统计的基本概念

为了更好地介绍数理统计所研究的问题, 先引入一些常用的概念和术语.

一、母体和子样

我们今后所讨论的统计问题主要属于下面这种类型: 从一个集合中选取一部分元素,对这部分元素的某些数量指标进行测量,根据测量获得的这些数据来推断这集合中全部元素的这些数量指标的分布情况. 在统计学中,我们把所研究的全部元素组成的集合称为母体, 或总体. 而把组成母体的每个元素称为个体. 例如在研究某批灯泡的平均寿命时,该批灯泡的全体就组成了母体,而

其中每个灯泡就是个体．在研究上海市男大学生的身高和体重的分布情况时，上海市的全体男大学生组成了母体，上海市的每个男大学生是个体．但是在统计里，由于我们关心的不是每个个体的种种具体特性，而仅仅是它的某一项或某几项数量指标 X（可以是向量）和该数量指标 X 在母体中的分布情况．在上述例子中 X 是表示灯泡的寿命或男大学生的身高和体重．就此数量指标 X 而言，每个个体所取的值是不同的．在试验中，抽取了若干个个体就观察到了 X 的这样或那样的数值，因而这个数量指标 X 是一个随机变量（或向量），而 X 的分布就完全描写了母体中我们所关心的那个数量指标的分布状况．由于我们关心的正是这个数量指标，因此我们以后就把母体和数量指标 X 可能取值的全体组成的集合等同起来．所谓母体的分布也就是指数量指标 X 的分布．

一般说来母体这个集合的大小与所研究的问题有关，如在灯泡的自动化生产中，为了保证正常生产需每隔一定时间抽取几个灯泡测试其寿命，并认为灯泡的寿命 X 服从正态分布．在这里母体被想象为所生产的灯泡寿命的可能取值的全体所组成的集合，这样母体被视为由无限个元素组成．并视 X 为连续型随机变量．

为了对母体的分布律进行各种研究，就必需对母体进行抽样观察．一般说来，我们还不止进行一次抽样观察，而要进行几次观察．通过观察就得到母体指标 X 的一组数值 $(x_1, x_2, \cdots, x_n)$，其中每个 x_i 是一次抽样观察的结果．即某一个被观察的个体的 X 指标值．$(x_1, \cdots, x_n)$ 称为容量为 n 的子样的观察值．由于我们是利用子样观察来对母体的分布进行推断，因而从母体中抽取子样进行观察时必须是随机的．直观地说，如果我们要研究上海市男大学生的身高的分布情况，那么在抽样时就希望上海市的每个男大学生具有同等的可能被抽到测量身高，因为只有这样才能经过多次观察比较全面地了解母体．所以对于随机抽样来说，对其某一

次观察结果而论，$(x_1, x_2, \cdots, x_n)$是完全确定的一组值，但它又是随每次抽样观察而改变的，由于我们要依据这一观察结果进行分析推断，并研究比较各种推断方法的好坏，因而一般考虑问题时，就不能把$(x_1, x_2, \cdots, x_n)$看为确定的数值，而应该看作为随机向量 $X=(X_1, X_2, \cdots, X_n)$①．称它为**容量是 n 的子样**，因而对子样也有分布可言．在不同的抽样观察中，X 得到不同的现实．$X=(X_1, X_2, \cdots, X_n)$ 所可能取值的全体(这儿是 n 维空间或其中的一个子集)称为**子样空间**．一个子样观察值$(x_1, x_2, \cdots, x_n)$就是子样空间中的一个点．

我们抽取子样的目的是为了对母体的分布律进行各种分析推断，因而要求抽取的子样能很好地反映母体的特性，这就必须对随机抽样的方法提出一定的要求．通常提出下面两点：(i) 代表性：要求子样的每个分量 X_i 与所考察的母体 $\mathbb{X}$ 具有相同的分布 $F(x)$；(ii) 独立性：$X_1, X_2, \cdots, X_n$ 为相互独立的随机变量，也就是说，每个观察结果既不影响其它观察结果，也不受其它观察结果的影响．(在有限母体中，子样的各个观察结果可以是不独立的．)

满足上述两点性质的子样称为**简单随机子样**，获得简单随机子样的抽样方法称为简单随机抽样．在今后，如果不作特殊声明，所说的子样将理解为简单随机子样，对于简单随机子样 $X=(X_1, X_2, \cdots, X_n)$，其分布可以由母体 $\mathbb{X}$ 的分布函数 $F(x)$ 完全决定，X的分布函数是 $\prod_{i=1}^{n} F(x_i)$．

[例 1] 某血吸虫防治所为了研究某地区血吸虫病的流行情况，对该地区所属的各生产队编制了一组卡片，每张卡片上标明一个生产队的血吸虫病的感染率．并将感染率的高低分为 4 类 (无患者，轻度流行，中度流行和重度流行)．分别用数字 0, 1, 2, 3 表

---

① 第二册中，除协方差矩阵符号用黑体外，其他矩阵、向量符号均不用黑体．

示这4类，那么在每张卡片上可以标出数字0到3中的一个，以表示该生产队血吸虫病的流行情况属于那一类．用$\mathbb{X}$记这数字，这$\mathbb{X}$就是我们要研究的指标．

现在，母体是全部卡片，设一共有N张．如果用返回抽取方法从这些卡片中任意抽取$n=10$张，即每抽一张卡片记录下该卡片上的感染率的类别数字后立即放回去，然后再抽下一张．在这里各次抽得的卡片其指标X取值k $(k=0, 1, 2, 3)$的概率不变，从抽得的10张卡片上观察到的随机变量$\mathbb{X}$的值$(x_1, x_2, \cdots, x_{10})$是一个10维随机向量的现实，这个10维随机向量的观察值构成一个子样；子样的容量$n=10$；子样空间是由所有可能的各组$(x_1, x_2, \cdots, x_{10})$组成．由于$x_k$ $(k=1, 2, \cdots, 10)$只可能取值0，1，2，3，因此现在的子样空间由4^{10}个点组成．

二、统计量和子样矩

子样是母体的代表和反映，但在我们抽取子样之后，并不直接利用子样进行推断．而需要对子样进行一番"加工"和"提炼"，把子样中所包含的关于我们所关心的事物的信息集中起来．这便是针对不同的问题构造出子样的某种函数．这种函数在统计学中称为统计量．严格地说，一个统计量就是观察的n维随机向量即子样$X=(X_1, X_2, \cdots, X_n)$的一个(波雷尔可测)函数，且要求它不包含有任何未知参数．因此它也是一个随机变量(或向量)．

若设X_1, X_2是从具有分布密度为$N(\mu, \sigma^2)$的正态母体中抽取的一个2维子样，其中μ, σ^2是未知参数．则$\frac{1}{2}(X_1+X_2)-\mu, \frac{X_1}{\sigma}$都不是统计量，因为它们含有未知参数．而$X_1$，$X_1+3$，$X_1^2+X_2^2$却都是统计量．

下面先介绍一些常用的统计量——子样矩．

设$(X_1, X_2, \cdots, X_n)$是从母体$\mathbb{X}$中抽取的一个子样，称统计量

$$\overline{X} \triangleq \frac{1}{n}\sum_{i=1}^{n} X_i \tag{1}$$

为子样均值；统计量

$$S_n^2 \triangleq \frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2 \tag{2}$$

为子样方差；统计量

$$A_\nu \triangleq \frac{1}{n}\sum_{i=1}^{n} X_i^\nu \tag{3}$$

为子样的 ν 阶矩（或 ν 阶原点矩）；统计量

$$B_\nu \triangleq \frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^\nu \tag{4}$$

为子样的 ν 阶中心矩.

现在我们研究子样矩的期望与方差，在下面的讨论中，我们用 μ 表示母体的均值，σ^2 表示母体的方差，α_k 表示母体的 k 阶原点矩，μ_k 表示母体的 k 阶中心矩，即记

$$\begin{aligned} &EX \triangleq \mu, D(X)=E(X-\mu)^2 \triangleq \sigma^2 \\ &EX^k \triangleq \alpha_k, E(X-\mu)^k \triangleq \mu_k \end{aligned} \tag{5}$$

并且约定，在我们用到 α_k（或 μ_k）时，假定它是存在的.

定理 1 设母体 X 服从分布 $F(x)$，$X=(X_1, \cdots, X_n)$ 是从该母体中抽得的一个简单随机子样，如果 $F(x)$ 的二阶矩存在，则对子样均值 $\overline{X}$，有

$$E\overline{X}=\mu$$

$$D(\overline{X})=\frac{\sigma^2}{n}$$

[证明] $E(\overline{X})=E\left(\frac{1}{n}\sum_{i=1}^{n}X_i\right)=\frac{1}{n}\sum_{i=1}^{n}E(X_i)=\frac{1}{n}\sum_{i=1}^{n}\mu=\mu$

$$D(\overline{X})=E(\overline{X}-\mu)^2=E\left(\frac{1}{n}\sum_{i=1}^{n}X_i-\mu\right)^2$$

$$=\frac{1}{n^2}E\left[\sum_{i=1}^{n}(X_i-\mu)\right]^2=\frac{1}{n^2}\sum_{i=1}^{n}E(X_i-\mu)^2=\frac{\sigma^2}{n}$$

定理2 如果 $F(x)$ 存在4阶矩,则对子样方差,有

$$E S_n^2 = \frac{n-1}{n}\sigma^2$$

$$D(S_n^2) = \frac{\mu_4-\mu_2^2}{n} - \frac{2(\mu_4-2\mu_2^2)}{n^2} + \frac{\mu_4-3\mu_2^2}{n^3} \quad (\mu_2=\sigma^2)$$

[证明] $E S_n^2 = E\left\{\frac{1}{n}\sum_{i=1}^n X_i^2 - \frac{1}{n^2}\left(\sum_{i=1}^n X_i\right)^2\right\}$

$$= \alpha_2 - \frac{1}{n^2} E\left\{\sum_{i=1}^n X_i^2 + \sum\sum_{i\neq j} X_i X_j\right\}$$

$$= \alpha_2 - \frac{1}{n^2}(n\alpha_2 + n(n-1)\mu^2)$$

$$= \frac{n-1}{n}(\alpha_2-\mu^2) = \frac{n-1}{n}\sigma^2$$

现记 $Y_i \triangleq X_i-\mu$,则 $EY_i=0$, $D(Y_i)=\sigma^2$, $EY_i^4=\mu_4$, 且有

$$n^2(S_n^2)^2 = \left[\sum_{i=1}^n (X_i-\overline{X})^2\right]^2 = \left[\sum_{i=1}^n (X_i-\mu)^2 - n(\overline{X}-\mu)^2\right]^2$$

$$= \left(\sum_{i=1}^n Y_i^2 - n\overline{Y}^2\right)^2 = \left(\sum_{i=1}^n Y_i^2\right)^2 - \frac{2}{n}\left(\sum_{i=1}^n Y_i^2\right)\left(\sum_{j=1}^n Y_j\right)^2$$

$$+ \frac{1}{n^2}\left(\sum_{j=1}^n Y_j\right)^4$$

$$= \left(\sum_{i=1}^n Y_i^4 + \sum\sum_{i\neq j} Y_i^2 Y_j^2\right) - \frac{2}{n}\left(\sum\sum_{i\neq j} Y_i^2 Y_j^2 + \sum_{i=1}^n Y_i^4\right.$$

$$\left.+ \sum_k \sum\sum_{i\neq j} Y_k^2 Y_i Y_j\right) + \frac{1}{n^2}\left(\sum_{j=1}^n Y_j^4 + 3\sum\sum_{i\neq j} Y_i^2 Y_j^2\right.$$

$$\left.+ 2\sum_k \sum\sum_{i\neq j} Y_k^2 Y_i Y_j + \sum' Y_i Y_j Y_k Y_l\right)^{①}$$

对上式两边取期望得:

$$n^2 E(S_n^2)^2 = n\mu_4 + n(n-1)\mu_2^2 - \frac{2}{n}[n(n-1)\mu_2^2$$

$$+ n\mu_4] + \frac{1}{n^2}[n\mu_4 + 3n(n-1)\mu_2^2]$$

$$= \left(n-2+\frac{1}{n}\right)\mu_4 + \left(n-2+\frac{3}{n}\right)(n-1)\mu_2^2$$

① Σ' 表示对 $i\neq k, j\neq l, (i,j)\neq(l,k)$ 的 i, j, k, l 求和。

所以

$$D(S_n^2)=E(S_n^2)^2-(ES_n^2)^2$$

$$=\frac{1}{n^2}\left[\left(n-2+\frac{1}{n}\right)\mu_4+\left(n-2+\frac{3}{n}\right)(n-1)\mu_2^2\right]$$

$$-\left(\frac{n-1}{n}\right)^2\mu_2^2$$

$$=\frac{1}{n^2}\left(n-2+\frac{1}{n}\right)\mu_4+(n-1)(3-n)\frac{\mu_2^2}{n^3}$$

$$=\frac{\mu_4-\mu_2^2}{n}-\frac{2(\mu_4-2\mu_2^2)}{n^2}+\frac{\mu_4-3\mu_2^2}{n^3}$$

定理3 如果 $F(x)$ 存在 2ν 阶矩，则对子样的 ν 阶原点矩 A_ν，有

$$EA_\nu=\alpha_\nu$$

$$D(A_\nu)=\frac{\alpha_{2\nu}-\alpha_\nu^2}{n}$$

[证明] $EA_\nu=E\left(\frac{1}{n}\sum_{i=1}^{n}X_i^\nu\right)=\frac{1}{n}\sum_{i=1}^{n}E(X_i^\nu)=\alpha_\nu$

$$D(A_\nu)=EA_\nu^2-(EA_\nu)^2=E\left(\frac{1}{n}\sum_{i=1}^{n}X_i^\nu\right)^2-\alpha_\nu^2$$

$$=E\left(\frac{1}{n^2}\sum_{i=1}^{n}X_i^{2\nu}+\frac{1}{n^2}\sum\sum_{i\neq j}X_i^\nu X_j^\nu\right)-\alpha_\nu^2$$

$$=\frac{1}{n}\alpha_{2\nu}+\frac{1}{n^2}\cdot n(n-1)\alpha_\nu^2-\alpha_\nu^2=\frac{\alpha_{2\nu}-\alpha_\nu^2}{n}.$$

三、顺序统计量和经验分布函数

设 $(X_1, \cdots, X_n)$ 是从母体 X 中抽取的一个子样. 记 $(x_1, x_2, \cdots, x_n)$ 是子样的一个观察值，将观察值的各分量按大小递增次序排列，得到

$$x_1^*\leqslant x_2^*\leqslant\cdots\leqslant x_n^*$$

当 $(X_1, \cdots, X_n)$ 取值为 $(x_1, \cdots, x_n)$ 时，我们定义 $X_k^{(n)}$ 取值为 x_k^*. 称由此得到的 $X_1^{(n)}, \cdots, X_n^{(n)}$ 为 $(X_1, \cdots, X_n)$ 的一组**顺序统计量**. 显然 $X_1^{(n)}\leqslant X_2^{(n)}\leqslant\cdots\leqslant X_n^{(n)}$, $X_1^{(n)}=\min\limits_{1\leqslant i\leqslant n}X_i$, 即 $X_1^{(n)}$ 的观察

值是子样观察值中最小的一个，而 $X_n^{(n)}=\max\limits_{1\leqslant i\leqslant n} X_i$，即 $X_n^{(n)}$ 的观察值是子样观察值中最大的一个．关于顺序统计量的进一步讨论可参阅第 7 章 § 2．

记

$$F_n^*(x)\triangleq\begin{cases}0, & \text{当 } x\leqslant x_1^* \\ \dfrac{k}{n}, & \text{当 } x_k^*<x\leqslant x_{k+1}^*, \quad k=1,2,\cdots,n-1 \\ 1, & \text{当 } x>x_n^*\end{cases}$$

显然 $0\leqslant F_n^*(x)\leqslant 1$，且作为 x 的函数是一非减左连续函数，把 $F_n^*(x)$ 看作为 x 的函数，它具备分布函数所要求的性质，故称为经验分布函数（或子样分布函数）．

经验分布函数也是子样的函数，它与子样矩之间具有下列关系：设 $(x_1, x_2, \cdots, x_n)$ 是子样观察值，$F_n^*(x)$ 是对应的经验分布函数，则有：

$$\bar{x}=\frac{1}{n}\sum_{i=1}^{n} x_i=\int x dF_n^*(x)$$

$$s_n^2=\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2=\int(x-\bar{x})^2 dF_n^*(x)$$

$$a_\nu=\frac{1}{n}\sum_{i=1}^{n} x_i^\nu=\int x^\nu dF_n^*(x), \quad \nu=2, 3, \cdots$$

$$b_\nu=\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^\nu=\int(x-\bar{x})^\nu dF_n^*(x), \quad \nu=3, 4, \cdots$$

§ 3．求估计量的两种常用方法

一、问题的提出

参数估计是统计推断的基本问题之一，这类问题对我们并不生疏．如在第一册例 1.3.7 中提出要求估计鱼池中鱼的条数；在 § 2 中提出要求估计上海市男大学生的平均身高等，这些都是参数估计问题．下面再举几个例子．

〔例 1〕 灯泡厂生产的灯泡，由于种种随机因素的影响，每批生产出来的灯泡，其中每个灯泡的使用寿命是不一致的，也就是说，灯泡的使用寿命 X 是一个随机变量．由中心极限定理和实际经验知道，灯泡使用寿命 X 服从正态 $N(\mu,\ \sigma^2)$ 分布．但一般我们事先并不能确切知道 $(\mu,\ \sigma^2)$ 的具体数值，而只知道它们是落在某个范围中，例如 $(0,\ \infty)\times(0,\ \infty)$ 或某个更小的区域里．为了断定所生产的这批灯泡的质量，自然提出要求估计这批灯泡的平均寿命以及寿命长短的差异程度．即要求估计 μ 和 σ^2．有时还希望以一定的可靠性来估计其平均寿命 μ 是在某个范围内或者不低于某个数．

〔例 2〕 大家知道，纺织厂细纱机上的断头次数可以用普阿松分布 $P(\lambda)$ 来描述，现在希望知道每只纱绽在某一时间间隔内断头数为 k 次的概率 $(k=0,\ 1,\ 2,\ 3,\ \cdots)$ 也就是要找出母体的分布律．而对于普阿松分布，只要知道数学期望 λ——平均断头次数，就可以确定其分布律 $P(\lambda)$．同样我们常常不知道 λ 的确切值，而只知道它落在某个范围里，例如 $(0,\ \infty)$ 或某个区间中．在这里，我们也提出了要求估计母体分布的均值的问题．

由上述可见，在参数估计问题中，我们总是首先假设母体 X 具有一族可能的分布 F，且 F 的函数形式是已知的，仅包含有几个未知参数．记 θ 是支配这分布的未知参数（可以是向量），在统计学上，我们把分布 F 的未知参数 θ 的全部可容许值组成的集合称为参数空间，记为 Θ．

今后，我们用 $F(\cdot;\theta)$ 表示 X 的分布，又称集合 $\{F(\cdot;\theta),\ \theta\in\Theta\}$ 为 X 的分布函数族．类似地，如果 X 是连续型随机变量，我们有概率密度函数族；如果 X 是离散型随机变量，我们有概率分布族．

在例 1 中，灯泡寿命 X 服从 $N(\mu,\ \sigma^2)$ 分布，参数空间是 $\Theta=\{(\mu,\ \sigma^2):\ 0<\mu<\infty,\ \sigma^2>0\}$．$\{N(\mu,\ \sigma^2),\ (\mu,\ \sigma^2)\in\Theta\}$ 就是

X 的分布函数族。

在例 2 中记断头次数为 X, X 服从 $P(\lambda)$，此处参数空间是 $\Theta=\{\lambda; 0<\lambda<\infty\}$, $\{P(\lambda); \lambda\in\Theta\}$ 是 X 的概率分布族.

一个参数估计问题就是要求通过子样估计母体分布所包含的未知参数 θ 或 θ 的某函数.

由于母体分布的函数形式是已知的，未知的仅是一个或几个未知参数，而母体的真分布也完全由这些参数所完全决定. 因此通过估计参数可以估计母体的真分布. 但我们有时也直接感兴趣于估计这些参数的值.

那末如何估计未知参数呢？仍以例 1 为例，为了估计灯泡的平均寿命，当然要抽取若干个灯泡作试验(即抽取子样)，若它们的寿命分别是 $X_1, X_2, \cdots, X_n$. 由大数定律知道，当 n 很大时，$\bar{X}=\frac{1}{n}(X_1+\cdots+X_n)$ 以很大的概率与 $\mu=EX$ 任意接近，因而自然地把子样寿命的平均值 $\bar{X}$ 作为母体平均寿命 μ 的估计量.

一般地，设母体具有分布族 $\{F(x;\theta), \theta\in\Theta\}$, $X_1, X_2, \cdots, X_n$ 是它的一个子样. 点估计问题就是要求构造一个统计量 $T(X_1,\cdots,X_n)$ 作为参数 θ 的估计(T 的维数与 θ 的维数相同). 在统计学上，我们称 T 为 θ 的估计量. 如果 $x_1, \cdots, x_n$ 是子样的一组观察值，代入统计量 T 就得到 T 的具体数值，这个数值常称为 θ 的估计值，今后，估计量和估计值这两个名词将不强调它们的区别通称为估计.

除了点估计问题以外，还有另一类估计问题，它要求用区间 (T_1, T_2) 作为 θ 可能取值范围的一种估计. 这个区间称为置信区间，这类估计问题称为区间估计问题. 由于参数的区间估计方法与假设检验具有密切的联系，所以放在 § 3.3 中介绍.

在这里，估计量 T 是子样 $X_1, \cdots, X_n$ 的函数，它不可以包含

有未知参数，也就是说 T 是一个作估计用的统计量．是随机变量（向量）．当我们获得了一组子样观察值 $(x_1, \cdots, x_n)$ 之后，就用 $T(x_1, \cdots, x_n)$ 作为未知参数 θ 的估计值．对于不同的子样观察值，所得的估计值是不同的．寻找一个估计量就是要找一个依据观察结果求得未知参数 θ 的估计值的方法．而不是具体去寻找一个估计值．衡量一个估计的好坏当然不能按一次试验结果得到的参数估计值与参数真值的偏差大小来确定，而必需从总体出发，也就是要以多次试验结果所得的估计值与参数真值的偏差大小来确定．这方面的详细讨论请阅第五章，这里仅介绍两种常用的求估计量的方法．

二、矩方法

矩方法是一种古老的估计方法．大家知道，矩是描写随机变量的最简单的数字特征．子样来自于母体，从定理 2.1~2.3 看到子样矩在一定程度上也反映了母体矩的特征，因而自然想到用子样矩作为母体矩的估计．

设 $\{F(x;\theta),\ \theta\in\Theta\}$ 是母体 X 的可能分布族，$\theta=(\theta_1, \cdots, \theta_k)$ 是待估计的未知参数．假定母体分布的 k 阶矩存在，则母体分布的 ν 阶矩

$$\alpha_\nu(\theta_1, \cdots, \theta_k)=\int_{-\infty}^{\infty} x^\nu dF(x;\theta),\quad 1\leqslant\nu\leqslant k$$

是 $\theta=(\theta_1, \cdots, \theta_k)$ 的函数．

对于子样 $X=(X_1, \cdots, X_n)$，其 ν 阶子样矩是

$$A_\nu=\frac{1}{n}\sum_{i=1}^{n} X_i^\nu,\quad 1\leqslant\nu\leqslant k$$

现在用子样矩作为母体矩的估计，即令

$$\alpha_\nu(\theta_1, \cdots, \theta_k)=A_\nu=\frac{1}{n}\sum_{i=1}^{n} X_i^\nu,\quad \nu=1, 2, \cdots, k \tag{1}$$

这样，(1)式确定了包含 k 个未知参数 $\theta=(\theta_1, \cdots, \theta_k)$ 的 k 个

方程式. 解此方程组 (1) 就可以得到 $\theta=(\theta_1, \cdots, \theta_k)$ 的一组解 $\hat{\theta}=(\hat{\theta}_1, \cdots, \hat{\theta}_k)$. 因为 A_v 是随机变量, 故解得的 $\hat{\theta}$ 也是随机变量. 现在将 $\hat{\theta}_1, \cdots, \hat{\theta}_k$ 分别作为 $\theta_1, \cdots, \theta_k$ 的估计, 称为**矩方法的估计**, 这种求估计量的方法称为**矩方法**.

[例 3] 母体均值和方差的矩估计.

设 $X_1, \cdots, X_n$ 是一子样, 设母体的二阶矩存在, 则有 $\alpha_2=\sigma^2+\mu^2$. 用矩方法得方程组

$$\begin{cases} \hat{\mu}=\dfrac{1}{n}\sum\limits_{i=1}^{n}X_i=\overline{X} \\ \hat{\mu}^2+\hat{\sigma}^2=\hat{\alpha}_2=\dfrac{1}{n}\sum\limits_{i=1}^{n}X_i^2 \end{cases}$$

解之得

$$\hat{\mu}=\overline{X}$$

$$\hat{\sigma}^2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2=S_n^2$$

所以母体均值 μ 和方差 σ^2 的矩估计分别是子样均值 $\overline{X}$ 和子样方差 S_n^2.

运用定理 2.1 有

$$E\hat{\mu}=E\overline{X}=\mu$$

$$D(\hat{\mu})=\frac{1}{n}\sigma^2$$

和

$$E(\hat{\sigma}^2)=\frac{n-1}{n}\sigma^2$$

由此可见, $\hat{\mu}$ 作为 μ 的估计它是在 μ 的真值的周围波动, 且其平均值恰好是真值 μ. 这一性质在统计学上称为**无偏性**.

一般地, 如果 $T(X)$ 是未知参数 θ 的一个估计量, 且满足下面的关系式,

$$E_\theta T(X)=\theta, \text{ 对一切 } \theta\in\Theta \tag{2}$$

则称 $T(X)$ 是 θ 的**无偏估计**.

由上可见, 子样均值 $\overline{X}$ 是母体均值 μ 的无偏估计, 同样如令

$$S_n^{*2} \triangleq \frac{n}{n-1} S_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2$$

由于
$$ES_n^{*2} = \frac{n}{n-1} ES_n^2 = \sigma^2$$

所以 S_n^{*2} 是母体方差 σ^2 的无偏估计.

[例 4] 设 $X=(X_1, \cdots, X_n)$ 是取自均匀分布 $\{f(x;\theta), \theta>0\}$

$$f(x;\theta)=\begin{cases}\dfrac{1}{\theta}, & \text{当 } 0<x<\theta \\ 0, & \text{其它}\end{cases}$$

的子样,试求 θ 的矩估计.

因为
$$\mu=\int_{-\infty}^{\infty} x f(x;\theta) dx = \frac{1}{\theta}\int_0^{\theta} x\, dx = \frac{\theta}{2}$$

由矩方法,令
$$\frac{\hat{\theta}}{2} = \hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i = \overline{X}$$

解之得 θ 的矩估计是
$$\hat{\theta} = 2\overline{X}$$

容易计算
$$E_\theta \hat{\theta} = \theta$$
$$D_\theta(\hat{\theta}) \doteq \frac{\theta^2}{3n}$$

这里 $\hat{\theta}=2\overline{X}$ 是 θ 的无偏估计.

三、最大似然估计方法

先考察两个简单的例子.

[例 5] 假设有某位同学与一位老战士一同进行实弹射击,两人同打一个靶子,每人各打一发,结果仅中一发,试问认为这一发是谁打中的较为合理?为什么?

[例 6] 假设在一个罐中放着许多白球和黑球,并假定已经知道两种球的数目之比是 1:3. 但不知道那种颜色的球多. 如果我们用返回抽样方法从罐中任意抽取 n 个球,则其中黑球的个数

为 x 的概率是

$$P(x;p)=\binom{n}{x}p^x q^{n-x}$$

其中 $q=1-p$, $p \triangleq \frac{\text{罐中黑球数目}}{\text{罐中全部球的数目}}$，由假定知道 p 仅可能取 $\frac{1}{4}$ 和 $\frac{3}{4}$ 两个值.

若 $n=3$, 则如何通过 x 来估计参数 p 呢？换句话说，在什么情况下取 $\hat{p}=\frac{1}{4}$, 而在另外的情况下取 $\hat{p}=\frac{3}{4}$ 呢？为此先计算一下抽样的可能结果 x 在这两种可能 p 值之下的概率(见下表)

x	0	1	2	3
$P\left(x;\frac{3}{4}\right)$	1/64	9/64	27/64	27/64
$P\left(x;\frac{1}{4}\right)$	27/64	27/64	9/64	1/64

由于子样来自于母体，因而子样应很好地反映母体的特征，而最大似然原理也就是以此为出发点. 在目前这个例子中，如果我们观察到的黑球数 $x=0$, 此时 $P\left(0;\frac{1}{4}\right)=27/64$, $P\left(0;\frac{3}{4}\right)=1/64$, 显然，$27/64 \gg 1/64$. 这说明，使 $x=0$ 的子样从带有 $p=\frac{1}{4}$ 的母体中抽取比从带有 $p=\frac{3}{4}$ 的母体中抽取更可能发生，因而取 $\frac{1}{4}$ 作为 p 的估计比取 $\frac{3}{4}$ 作为 p 的估计更合理. 类似地 $x=1$ 时也取 $\frac{1}{4}$ 作为 p 的估计，而当 $x=2, 3$ 时取 $\frac{3}{4}$ 作为 p 的估计. 即定义估计量如下：

$$\hat{p}(x)=\begin{cases}\frac{1}{4}, & \text{当 } x=0,\ 1\\ \frac{3}{4}, & \text{当 } x=2,\ 3\end{cases}$$

它是对每个x，选取$\hat{p}(x)$使得

$$P(x;\hat{p}(x))\geqslant P(x;p')$$

其中p'是不同于$\hat{p}(x)$的另一值．这就是最大似然原理的基本思想．

一般地，设母体X具有分布密度族$\{f(x;\ \theta),\ \theta\in\Theta\}$，其中$\theta=(\theta_1,\ \theta_2,\ \cdots,\ \theta_k)$是一个未知的$k$维参数向量，需待估计．又设$(x_1,\cdots,x_n)$是子样$(X_1,\ \cdots,\ X_n)$的一个观察值，那么子样$(X_1,\cdots,X_n)$落在点$(x_1,\ \cdots,\ x_n)$的邻域里的概率是$\prod_{i=1}^{n}f(x_i;\theta)dx_i$．由此可见，$\theta$的变化会影响到$\prod_{i=1}^{n}f(x_i;\theta)dx_i$大小的变化．也就是说，概率$\prod_{i=1}^{n}f(x_i;\ \theta)dx_i$是$\theta$的函数．最大似然方法就是选取使得子样落在观察值$(x_1,\ \cdots,\ x_n)$的邻域里的概率$\prod_{i=1}^{n}f(x_i;\ \theta)dx_i$达到最大的参数值$\hat{\theta}$作为$\theta$的估计值．即对固定的$(x_1,\ \cdots,\ x_n)$，选取$\hat{\theta}$使得

$$\prod_{i=1}^{n}f(x_i;\hat{\theta})=\max_{\theta\in\Theta}\prod_{i=1}^{n}f(x_i;\theta)$$

从直观上讲，我们既然在一次试验中得到了观察$(x_1,\ \cdots,\ x_n)$，那么我们认为子样落在该观察值$(x_1,\ \cdots,\ x_n)$的邻域里这一事件是较易发生的，具有较大的概率．所以就应选取使这一概率达到最大的参数值作为真参数值的估计．当X是离散型变量时，上面的$f(x_i;\theta)$就取为概率分布，今后恒作此约定．为方便起见，记

$$L(x;\theta)\triangleq\prod_{i=1}^{n}f(x_i;\theta) \tag{3}$$

(θ可以是向量)它看作为θ的函数称为θ的**似然函数**．

如果选取使下式

$$L(x_1,\ \cdots,\ x_n;\hat{\theta})=\sup_{\theta\in\Theta}L(x_1,\ \cdots,\ x_n;\theta) \tag{4}$$

成立的$\hat{\theta}(X)=(\hat{\theta}_1(X),\ \cdots,\ \hat{\theta}_k(X))$作为$\theta$的估计，则称$\hat{\theta}(X)$是$\theta$的**最大似然估计**．

由于 $\log x$ 是 x 的单调函数，所以(4)式可等价地写为：

$$\log L(x_1, \cdots, x_n; \hat{\theta}) = \sup_{\theta \in \Theta} \log L(x_1, \cdots, x_n; \theta) \tag{5}$$

如果 Θ 是开集，且 $f(x; \theta)$ 关于 θ 可微，则满足(4)式的解 $\hat{\theta}$ 也一定满足下列似然方程

$$\left.\frac{\partial \log L(x_1, \cdots, x_n; \theta)}{\partial \theta_j}\right|_{\theta=\hat{\theta}} = 0, \quad j=1, \cdots, k \tag{6}$$

[例 7] 设 $X=(X_1, \cdots, X_n)$ 是取自正态母体 $N(\mu, \sigma^2)$ 的子样，其中 μ, σ^2 是未知参数，$-\infty<\mu<\infty$, $\sigma^2>0$. 试求 μ 和 σ^2 的最大似然估计.

显然

$$L(x; \mu, \sigma^2) = \frac{1}{(\sqrt{2\pi}\,\sigma)^n} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right\}$$

似然方程是

$$\begin{cases} \dfrac{\partial \log L}{\partial \mu} = \dfrac{1}{\sigma^2}\displaystyle\sum_{i=1}^{n}(x_i-\mu) = 0 \\ \dfrac{\partial \log L}{\partial \sigma^2} = \dfrac{1}{2\sigma^4}\displaystyle\sum_{i=1}^{n}(x_i-\mu)^2 - \dfrac{n}{2\sigma^2} = 0 \end{cases}$$

解之得：

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}$$

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2 = S_n^2$$

容易验证 $\hat{\mu}$, $\hat{\sigma}^2$ 满足关系式

$$L(x; \hat{\mu}, \hat{\sigma}^2) = \sup_{\substack{-\infty<\mu<\infty \\ \sigma^2>0}} L(x; \mu, \sigma^2)$$

所以 $\bar{X}$ 和 S_n^2 是 μ 和 σ^2 的最大似然估计.

[例 8] 设 $X=(X_1, \cdots, X_n)$ 是取自均匀分布

$$f(x; \theta) = \begin{cases} \dfrac{1}{\theta}, & 0<x<\theta \\ 0, & \text{其它} \end{cases} \qquad (\theta>0)$$

的子样，试求 θ 的最大似然估计.

此时　$L(x;\theta)=\prod_{i=1}^{n}f(x_i;\theta)=\begin{cases}\theta^{-n}, & 当\ 0\leqslant \max\limits_{1\leqslant i\leqslant n}x_i\leqslant\theta\\ 0, & 其它\end{cases}$

显然当 $\theta=\max\limits_{1\leqslant i\leqslant n}x_i$ 时，$L(x;\theta)$ 取到最大值，所以 $\hat{\theta}_L=\max\limits_{1\leqslant i\leqslant n}X_i=X_n^{(n)}$ 是 θ 的最大似然估计.

直接计算可得

$$E_\theta(\hat{\theta}_L)=\frac{n}{n+1}\theta$$

$$D_\theta(\hat{\theta}_L)=\frac{n\theta^2}{(n+1)^2(n+2)}$$

比较例 3 和例 7 可见，对于正态母体，μ 和 σ^2 的矩估计和最大似然估计是相同的，分别是子样均值和子样方差. 而对于具有均匀分布的母体，由例 4 和例 8 看到，未知参数 θ 的矩估计是

$$\hat{\theta}=2\overline{X}$$

最大似然估计是

$$\hat{\theta}_L=\max_{1\leqslant i\leqslant n}X_i$$

这里 $\hat{\theta}_L$ 不是 θ 的无偏估计，但 $\hat{\theta}_L$ 的方差比 $\hat{\theta}$ 的方差要小，当子样容量 n 很大时尤为显著. 因为

$$D_\theta(\hat{\theta}_L)=\frac{n\theta^2}{(n+1)^2(n+2)}<\frac{\theta^2}{3n}=D_\theta(\hat{\theta})$$

和　$$\lim_{n\to\infty}\frac{D_\theta(\hat{\theta}_L)}{D_\theta(\hat{\theta})}=\lim_{n\to\infty}\frac{3n^2}{(n+1)^2(n+2)}=0$$

又因为　$$\lim_{n\to\infty}E_\theta(\hat{\theta}_L)=\lim_{n\to\infty}\frac{n}{n+1}\theta=\theta$$

所以直观上很明显，当 n 较大时，$\hat{\theta}_L$ 比 $\hat{\theta}$ 更集中在真参数值 θ 的周围波动. 因此作为 θ 的估计量，$\hat{\theta}_L$ 比 $\hat{\theta}$ 好.

一般地若 T_1 是 θ 的一个无偏估计，关于 θ 的任一无偏估计 T_2 成立下式

$$D_\theta(T_1)\leqslant D_\theta(T_2),\ 对一切\ \theta\in\Theta \tag{7}$$

则称 T_1 是 θ 的**最小方差无偏估计**.

习　题

1. 设母体 X 服从贝努里 $b(1, p)$ 分布. $(P(X=1)=p)$. 其中 p 是未知参数. $X=(X_1, \cdots, X_5)$ 是从中抽取的一简单随机子样.

(1) 写出它的子样空间和 X 的概率分布;

(2) 指出 X_1+X_2, $\min_{1\le i\le 5} X_i$, X_5+2p, $(X_5-X_1)^2$ 之中, 那些是统计量, 那些不是统计量, 为什么?

(3) 如果 X 的一个观察值是 (0, 1, 0, 1, 1), 计算它的子样平均值和子样方差.

2. 设母体 X 服从正态 $N(\mu, \sigma^2)$ 分布, 其中 μ 是已知的而 σ^2 是未知的. 设 $X=(X_1, X_2, X_3)$ 是从中抽取的一个简单子样.

(1) 写出它的子样空间和 X 的分布密度;

(2) 指出 $X_1+X_2+X_3$, $X_2+2\mu$, $\max(X_1, X_2, X_3)$, $\sum_{i=1}^{3}\frac{X_i^2}{\sigma^2}$, $\frac{X_3-X_1}{2}$ 之中, 那些是统计量, 那些不是统计量, 为什么?

3. 设母体 X 服从 $b(1, p)$ 分布, $X=(X_1, \cdots, X_n)$ 是一子样, 试求子样均值 $\overline{X}$ 的数学期望 $E\overline{X}$ 和方差 $D(\overline{X})$ 及子样方差的数学期望 $E(S_n^2)$.

4. 设母体服从普阿松分布 $P(\lambda)$, $X=(X_1, \cdots, X_n)$ 是一子样. 试

(1) 写出 X 的概率分布;

(2) 计算 $E\overline{X}$, $D(\overline{X})$ 和 $E(S_n^2)$.

5. 设母体 X 服从贝努里分布 $b(1, p)$, 其中 p 是未知参数, $0<p<1$. $(X_1, \cdots, X_n)$ 是从该母体中抽取的子样. 试求未知参数 p 的矩估计和最大似然估计.

6. 设母体 X 服从普阿松分布 $P(\lambda)$, $\lambda>0$ 是未知参数. $X=(X_1, \cdots, X_n)$ 是一简单随机子样. 试求未知参数 λ 的矩估计和最大似然估计.

7. 设母体 X 服从均匀分布, 其密度函数

$$f(x;\theta)=\begin{cases}1, & \theta\in\left[\theta-\frac{1}{2}, \theta+\frac{1}{2}\right]\\ 0, & \text{其它}\end{cases}$$

θ 是未知参数, $X=(X_1, \cdots, X_n)$ 是一子样, 试求 θ 的最大似然估计.

8. 设母体 X 具有分布密度

$$f(x;\alpha)=(\alpha+1)x^{\alpha},\ 0<x<1$$

其中$\alpha>-1$，是未知参数．$X=(X_1, \cdots, X_n)$是一子样，试求参数α的最大似然估计和矩估计．

9．设母体X服从对数正态分布，其分布密度是

$$f(x;\mu,\sigma^2)=(2\pi\sigma^2)^{-\frac{1}{2}}x^{-1}\exp\left\{-\frac{1}{2\sigma^2}(\log x-\mu)^2\right\},\quad x>0$$

其中$-\infty<\mu<\infty$，$\sigma^2>0$是未知参数，$X=(X_1, \cdots, X_n)$是一子样，试求μ和σ^2的矩估计．

10．设母体X服从二项分布$b(k,p)$，k是正整数，$0<p<1$，两者都是未知参数．$X=(X_1, \cdots, X_n)$是一子样，试求k和p的矩估计．

第二章　抽样分布

在第一章中已经看到，如要进一步分析一个估计量的好坏，就需知道估计量的分布，至少要知道估计量的某些数字特征．由于估计量是用于估计问题的统计量，统计量是我们对母体的分布律或数字特征进行推断的基础．因此求统计量的分布是数理统计的基本问题之一．

我们所感到兴趣的是下面两类问题．第一类问题是：对于任意一个自然数 n，要找出给定的统计量 $U_n=f(X_1, X_2, \cdots, X_n)$ 的分布，这分布称为这统计量的**精确分布**．求统计量的精确分布对于数理统计中的所谓**小样问题**(即子样容量比较小时的统计问题)的研究是非常有用的．

第二类问题是：不对任何个别的 n 求出统计量 U_n 的分布，而只求出当 $n\to\infty$ 时，统计量 U_n 的极限分布，这极限分布对于数理统计中的所谓**大样问题**(即子样容量较大时的统计问题)的研究很有用处．

一般说来，要确定一个统计量的精确分布是非常复杂的，可是对于一些重要的特殊情形，如正态母体，这个问题有较简单的解法．

在今后各章中将会看到，正态母体的研究处于特别显著的地位．这一方面是由于其统计量的精确分布的数学分析比较容易；另一重要原因是：在许多领域的统计研究中所遇到的母体，正态分布是它的一个很好的近似．当然，中心极限定理也保证了这一状况．这章内容是为研究正态母体的统计量的精确分布服务的．

§1. 正态母体子样的线性函数的分布

在第一册 4.5 多元正态分布这一节中详细地讨论了多元正态分布的性质. 对于正态母体的子样 $X_1, \cdots, X_n$, 由于它们是独立同服从于正态 $N(\mu, \sigma^2)$分布的随机变量,它是多元正态分布的一种特殊情况. 因此这节的结论是第一册 4.5 中的有关结论在这一特殊条件下的推论,故这里只给出结论,不加证明.

设母体 $\mathbb{X}$ 服从正态分布 $N(\mu, \sigma^2)$,它的密度函数是

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}\exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\} \tag{1}$$

其中 μ 和 σ^2 分别是 $\mathbb{X}$ 的数学期望和方差.

设 $X_1, \cdots, X_n$ 是从该母体中抽得的一个容量为 n 的简单随机子样. 考察统计量

$$U=a_1X_1+\cdots+a_nX_n$$

其中 $a_1, \cdots, a_n$ 是已知常数. 它是子样的线性函数. 对它有

定理 1 设 $X_1, \cdots, X_n$ 是抽自正态母体 $N(\mu, \sigma^2)$ 的一个子样,统计量 U 是子样的任一确定的线性函数

$$U=a_1X_1+a_2X_2+\cdots+a_nX_n \tag{2}$$

则 U 也是正态随机变量,均值、方差分别为

$$E(U)=\mu\sum_{k=1}^{n}a_k \tag{3}$$

$$D(U)=\sigma^2\sum_{k=1}^{n}a_k^2 \tag{4}$$

在(2)式中,特别地取 $a_k=\frac{1}{n}, k=1, \cdots, n$,此时得到的$U$是子样均值 $\overline{X}$.

系 1 设 $X_1, X_2, \cdots, X_n$ 是从正态母体 $N(\mu, \sigma^2)$所抽得的一个子样,则子样均值

$$\overline{X}=\frac{1}{n}\sum_{k=1}^{n}X_k \tag{5}$$

也是正态随机变量，且

$$E(\overline{X})=\mu \tag{6}$$

$$D(\overline{X})=\frac{1}{n}\sigma^2 \tag{7}$$

由此可见，$\overline{X}$ 具有与 X 相同的均值，但是它更向数学期望集中，集中程度与子样容量 n 的大小有关。

定理2 设

(1) $X_1, X_2, \cdots, X_n$ 是独立同分布随机变量，同服从于正态分布 $N(\mu, \sigma^2)$；

(2) $A=(a_{ij})$ 是一 $p\times n$ 矩阵，记

$$Y\triangleq\begin{pmatrix}Y_1\\Y_2\\\vdots\\Y_p\end{pmatrix}=\begin{pmatrix}a_{11}\cdots a_{1n}\\a_{21}\cdots a_{2n}\\\cdots\cdots\cdots\\a_{p1}\cdots a_{pn}\end{pmatrix}\begin{pmatrix}X_1\\X_2\\\vdots\\X_n\end{pmatrix}\triangleq AX \tag{8}$$

$$X\triangleq(X_1X_2\cdots X_n)^T$$

则 $Y_1, \cdots, Y_p$ 也是正态随机变量，均值、方差、协方差分别为：

$$\begin{gathered}EY_i=\mu\sum_{k=1}^{n}a_{ik},\ i=1,\ \cdots,\ p\\ D(Y_i)=\sigma^2\sum_{k=1}^{n}a_{ik}^2,\ i=1,\ \cdots,\ p\\ \operatorname{cov}(Y_i,\ Y_j)=\sigma^2\sum_{k=1}^{n}a_{ik}a_{jk},\ i,\ j=1,\ \cdots,\ p\end{gathered} \tag{9}$$

特别地，当 $\mu=0$，且 A 是一 $n\times n$ 正交阵时，$Y_1, Y_2, \cdots, Y_n$ 也是相互独立且同服从于 $N(0, \sigma^2)$ 分布的随机变量。

由此可见，独立、均值为 0 的正态变量 $X_1,\cdots,X_n$ 通过正交变换所获得的新的随机变量 $Y_1,Y_2,\cdots,Y_n$ 也是独立、均值为 0 的正态随机变量，即 Y 与 X 的分布律相同。这反映了正态变量在正交变换下的某种不变性质。

§2. χ^2-分 布

一、χ^2-分布

定义 1 设 $X_1, X_2, \cdots, X_n$ 是相互独立，且同服从于 $N(0,1)$ 分布的随机变量，则称随机变量

$$\chi_n^2 = \sum_{i=1}^{n} X_i^2 \tag{1}$$

所服从的分布为 χ^2-分布.

定理 1 由(1)式定义的随机变量 χ_n^2 的分布密度函数是

$$\chi^2(x; n) = \begin{cases} \dfrac{1}{2^{\frac{n}{2}} \Gamma\left(\dfrac{n}{2}\right)} e^{-\frac{x}{2}} x^{\frac{n}{2}-1}, & \text{当 } x>0 \\ 0, & \text{当 } x \leqslant 0 \end{cases} \tag{2}$$

其中参数 n 称为自由度，它表示(1)中独立变量的个数.

［证明］ 先计算 χ_n^2 的分布函数 $F(x)$,

$$F(x) \triangleq P\{\chi_n^2 < x\} = \begin{cases} 0, & \text{当 } x \leqslant 0 \\ P\left\{\sum\limits_{i=1}^{n} X_i^2 < x\right\}, & \text{当 } x>0 \end{cases} \tag{3}$$

因为 $X_1, \cdots, X_n$ 相互独立，且同服从于 $N(0, 1)$分布，所以 $X_1, X_2, \cdots, X_n$ 的联合分布密度函数是

$$f_n(x_1, \cdots, x_n) = \frac{1}{(2\pi)^{\frac{n}{2}}} \exp\left\{-\frac{1}{2} \sum_{i=1}^{n} x_i^2\right\}$$

所以，对 $x>0$,

$$\begin{aligned} F(x) &= P\left\{\sum_{i=1}^{n} X_i^2 < x\right\} \\ &= \frac{1}{(2\pi)^{n/2}} \underset{\sum\limits_{i=1}^{n} x_i^2 < x}{\int \cdots \int} \exp\left\{-\frac{1}{2} \sum_{i=1}^{n} x_i^2\right\} dx_1 \cdots dx_n \end{aligned} \tag{4}$$

为了计算这个积分，作变换

$$\begin{cases} x_1 = \rho\cos\theta_1\cos\theta_2\cdots\cos\theta_{n-1} \\ x_2 = \rho\cos\theta_1\cos\theta_2\cdots\sin\theta_{n-1} \\ \quad\cdots \\ x_n = \rho\sin\theta_1 \end{cases}$$

则此变换的雅可比行列式为

$$J = \frac{\partial(x_1\cdots x_n)}{\partial(\rho,\ \theta_1,\ \cdots,\ \theta_{n-1})} = \rho^{n-1}D(\theta_1,\ \cdots,\ \theta_{n-1})$$

其中 $D(\theta_1,\ \cdots,\ \theta_{n-1})$ 是 $\theta_1,\ \cdots,\ \theta_{n-1}$ 的函数，不包含变量 ρ. 代入(4)式得

$$F(x) = \frac{1}{(2\pi)^{n/2}}\int_0^{\sqrt{x}}\int_{-\pi/2}^{\pi/2}\cdots\int_{-\pi/2}^{\pi/2}\int_{-\pi}^{\pi} e^{-\frac{1}{2}\rho^2}\ \rho^{n-1}d\rho D(\theta_1,\ \cdots,\ \theta_{n-1})$$
$$\times d\theta_1\cdots d\theta_{n-1} = C_n\int_0^{\sqrt{x}} e^{-\frac{1}{2}\rho^2}\rho^{n-1}d\rho \tag{5}$$

其中 $C_n = \dfrac{1}{(2\pi)^{n/2}}\displaystyle\int_{-\pi/2}^{\pi/2}\cdots\int_{-\pi/2}^{\pi/2}\int_{-\pi}^{\pi} D(\theta_1,\cdots,\theta_{n-1})d\theta_1\cdots d\theta_{n-1}$

令 $y \triangleq \rho^2$

则 $\rho = \sqrt{y},\qquad d\rho = \dfrac{1}{2\sqrt{y}}dy$

代入(5)式得

$$F(x) = C_n\cdot\frac{1}{2}\int_0^x e^{-\frac{y}{2}}y^{\frac{n}{2}-1}dy \tag{6}$$

因为 $F(+\infty) = 1 = \dfrac{C_n}{2}\displaystyle\int_0^\infty e^{-\frac{y}{2}}y^{\frac{n}{2}-1}dy$

利用等式 $\Gamma\left(\dfrac{n}{2}\right) = \displaystyle\int_0^\infty e^{-y}y^{\frac{n}{2}-1}dy$

容易算得 $C_n = \dfrac{1}{2^{\frac{n}{2}-1}\Gamma\left(\frac{n}{2}\right)}$

代入(6)式得到

$$F(x) = \frac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)}\int_0^x e^{-\frac{y}{2}}y^{\frac{n}{2}-1}dy \tag{7}$$

上式两边对 x 微分即得所证.

今后,如果随机变量 X 具有自由度为 n 的 χ^2-分布,将简单地写为 $X\sim\chi^2(n)$. 在图 1 中给出了当 $n=1, 4, 10, 20$ 时, χ^2-分布的密度函数曲线.

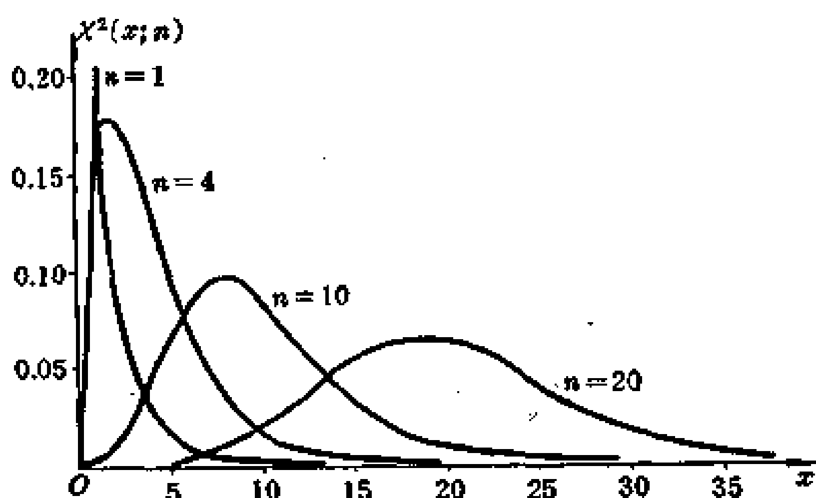


图 1　χ^2-分布的密度函数

[例 1]　设母体 X 服从正态分布 $N(\mu, \sigma^2)$, μ 是已知常数. $X_1, \cdots, X_n$ 是从该母体中抽取的一个容量为 n 的简单随机子样. 求统计量

$$\chi^2=\sum_{k=1}^{n}(X_k-\mu)^2 \tag{8}$$

的分布. 作变换

$$Y_k\triangleq\frac{X_k-\mu}{\sigma},\qquad k=1, \cdots, n$$

显然 $Y_1, Y_2, \cdots, Y_n$ 相互独立, 且同服从于 $N(0, 1)$分布的随机变量,因此由定理 1 知道

$$\frac{\chi^2}{\sigma^2}=\sum_{k=1}^{n}\left(\frac{X_k-\mu}{\sigma}\right)^2=\sum_{k=1}^{n}Y_k^2$$

服从于自由度为 n 的 χ^2-分布, 由此容易计算得到统计量(8)式的分布密度是

$$f(x)=\begin{cases}\dfrac{1}{2^{\frac{n}{2}}\sigma^n\Gamma\left(\dfrac{n}{2}\right)}e^{-\frac{x}{2\sigma^2}}x^{\frac{n}{2}-1}, & 当\ x>0\\ 0, & 当\ x\leqslant 0\end{cases}\tag{9}$$

由(8)定义的统计量通常也叫做 χ^2 统计量. 如果进一步我们用 $x_1, \cdots, x_n$ 表示子样的观察值,则

$$\sum_{k=1}^{n}(x_k-\mu)^2$$

是 χ^2 统计量的观察值.

定理 2 设 $X\sim\chi^2(n)$,则

(1) X 的特征函数是

$$\varphi(t)=(1-2it)^{-\frac{n}{2}}\tag{10}$$

(2) X 的数学期望和方差是

$$EX=n$$

和

$$D(X)=2n\tag{11}$$

[证明] (1) 由特征函数的定义得

$$\varphi(t)=\int_0^\infty e^{itx}\cdot\frac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)}e^{-\frac{x}{2}}x^{\frac{n}{2}-1}dx$$

$$=\frac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)}\int_0^\infty e^{-\left(\frac{1}{2}-it\right)x}x^{\frac{n}{2}-1}dx$$

$$=\frac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)}\cdot\frac{\Gamma\left(\frac{n}{2}\right)}{\left(\frac{1}{2}-it\right)^{\frac{n}{2}}}=(1-2it)^{-\frac{n}{2}}$$

(2)

$$EX=\varphi'(0)/i=n$$

$$EX^2=\varphi^{(2)}(0)/i^2=n(n+2)$$

$$DX^2=EX^2-(EX)^2=2n$$

定理 3 设 $X_1\sim\chi^2(n_1)$ 和 $X_2\sim\chi^2(n_2)$,且 X_1, X_2 相互独立,

则 $X_1+X_2\sim\chi^2(n_1+n_2)$.

[证明] 由定理 2, X_1 和 X_2 的特征函数分别是

$$\varphi_1(t)=(1-2it)^{-n_1/2}$$

和

$$\varphi_2(t)=(1-2it)^{-n_2/2}$$

因为 X_1 和 X_2 独立, 所以 X_1+X_2 的特征函数是

$$\varphi(t)=\varphi_1(t)\cdot\varphi_2(t)=(1-2it)^{-(n_1+n_2)/2}$$

即 X_1+X_2 也服从于 χ^2-分布, 自由度是 n_1+n_2.

系 1 设 $X_i\sim\chi^2(n_i)$, $i=1, \cdots, k$, 且相互独立.

则

$$\sum_{i=1}^{k}X_i\sim\chi^2\left(\sum_{i=1}^{n}n_i\right)$$

下面我们讨论一个比定理 3 更为深刻的结论

定理 4 (柯赫伦定理)

设 $X_1, \cdots, X_n$ 是相互独立同分布的随机变量, 且同服从于 $N(0, 1)$分布. 又设

$$Q_1+\cdots+Q_k=\sum_{i=1}^{n}X_i^2 \tag{12}$$

其中 Q_i $(i=1, \cdots, k)$ 是秩为 n_i 的 $(X_1, \cdots, X_n)$ 的非负二次型. 则 $Q_i, i=1, \cdots, k$ 相互独立, 且分别服从于自由度为 n_i 的 χ^2-分布的充要条件是

$$n_1+n_2+\cdots+n_k=n$$

[证明] 必要性是系 1 的直接结论. 现证充分性.

假设 $\sum_{i=1}^{k}n_i=n$ 成立, 由于 Q_i 的秩为 n_i, 由线性代数上的知识知道, 此时存在线性变换

$$Y_j=\sum_{l=1}^{n}p_{jl}X_l,\ j=n_1+\cdots+n_{i-1}+1, \cdots, n_1+\cdots+n_{i-1}+n_i$$

使得

$$Q_i=\sum_{j=n_1+\cdots+n_{i-1}+1}^{n_1+\cdots+n_{i-1}+n_i}Y_j^2 \tag{13}$$

记
$$X \triangleq \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}, \quad Y \triangleq \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}$$

$$P \triangleq \begin{pmatrix} p_{11} & \cdots & p_{1n} \\ \cdots & \cdots & \cdots \\ p_{n1} & \cdots & p_{nn} \end{pmatrix}$$

则有

$$Y = PX \tag{14}$$

由于
$$\sum_{i=1}^{n} X_i^2 = Q_1 + \cdots + Q_k = \sum_{j=1}^{n} Y_j^2$$

即
$$X^\tau X = Y^\tau Y = X^\tau P^\tau P X$$

由此得到

$$P^\tau P = I \tag{15}$$

(15) 式表明矩阵 P 是一个 $n \times n$ 正交阵. 运用定理 1.2 知道 $Y_1, \cdots, Y_n$ 是独立, 同服从于 $N(0,1)$ 的正态变量. 因此由(13)式得到: Q_i, $i=1, \cdots, k$ 相互独立, 且 $Q_i \sim \chi^2(n_i)$.

柯赫伦定理在 § 4.3 方差分析的研究中起重要作用.

二、非中心 χ^2-分布

定义 2 设 $X_1, X_2, \cdots, X_n$ 是独立正态随机变量, $EX_i = \mu_i$, $D(X_i) = \sigma^2$, $i = 1, 2, \cdots, n$, 且 μ_i 不全为零. 记

$$Y \triangleq \sum_{i=1}^{n} X_i^2 / \sigma^2 \tag{16}$$

则称 Y 是一非中心 χ^2 随机变量, 称它的分布为非中心 χ^2- 分布, 具有自由度 n 和非中心参数 $\sum\limits_{i=1}^{n} \mu_i^2/\sigma^2$. 并记为 $Y \sim \chi^2(n, \delta)$, 其中 $\delta = \sum\limits_{i=1}^{n} \mu_i^2/\sigma^2$. 显然 δ 是在(16)式中用 X_i 的数学期望代替 X_i 计算获得的值.

$\chi^2(n, \delta)$ 变量的分布密度函数是

$$\chi^2(x; n, \delta) = e^{-\delta/2}\sum_{m=0}^{\infty}\frac{\left(\frac{\delta}{2}\right)^m}{m!}\chi^2(x; 2m+n) \tag{17}$$

其中 $\chi^2(x;\ 2m+n)$ 是由(2)式所决定的自由度为 $2m+n$ 的 χ^2-分布的密度函数．由(17)式可见，非中心 χ^2-分布密度包含有 δ 和 n 两个参数．如果置 $\delta=0$，则 (17) 式就简化为由 (2) 式给出的 $\chi^2(n)$ 分布，因此有时又称由(2)式给出的 χ^2-分布为中心 χ^2-分布．

$\chi^2(n, \delta)$ 变量的特征函数是

$$\varphi(t) = (1-2it)^{-\frac{n}{2}} e^{\frac{i\delta t}{1-2it}} \tag{18}$$

容易计算 $\chi^2(n, \delta)$ 变量 X 的数学期望和方差分别是

$$EX = n+\delta$$
$$D(X) = 2n+4\delta \tag{19}$$

类似于系 1 也有：

如果 $X_1, X_2, \cdots, X_k$ 相互独立，且 $X_i \sim \chi^2(n_i, \delta_i)$, $i=1, 2, \cdots, k$.

则
$$\sum_{i=1}^{k} X_i \sim \chi^2\left(\sum_{i=1}^{k} n_i, \sum_{i=1}^{k}\delta_i\right)$$

§ 3. t-分布和 F-分布

一、t-分布

定义 1 设 $X\sim N(0, 1)$ 和 $Y\sim\chi^2(n)$，且 X 和 Y 相互独立，则称随机变量

$$T = \frac{X}{\sqrt{Y/n}} \tag{1}$$

所服从的分布为 t-分布．n 称为它的自由度，且记 $T\sim t(n)$．

在下节中将会看到，它在正态母体的抽样中是很自然地出现的．

定理 1 由(1)式所定义的 T 的分布密度函数是

$$t(x;n)=\frac{\Gamma[(n+1)/2]}{\Gamma(n/2)\sqrt{n\pi}}(1+x^2/n)^{-(n+1)/2} \tag{2}$$

[证明] (X, Y)的联合分布密度函数是

$$c\cdot e^{-x^2/2}e^{-y/2}y^{(n/2)-1} \tag{3}$$

作坐标变换:

$$x=r\sin\theta,\ y=r^2\cos^2\theta,\ 0<r<\infty,\ -\frac{\pi}{2}<\theta<\frac{\pi}{2}$$

则
$$dxdy=2r^2\cos\theta\, dr d\theta$$

记 R, Θ 是由关系式 $X=R\sin\Theta$, $Y=R\cos^2\Theta$, $0<R<\infty$, $-\frac{\pi}{2}<\Theta<\frac{\pi}{2}$,所决定的随机变量,则$(R, \Theta)$的联合密度函数是

$$2ce^{-r^2/2}r^n(\cos\theta)^{n-1} \tag{4}$$

由此可见,R 和 Θ 是独立的,Θ 的分布密度函数是

$$c_1(\cos\theta)^{n-1} \tag{5}$$

其中
$$c_1^{-1}=\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}}(\cos\theta)^{n-1}d\theta=\frac{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n+1}{2}\right)}$$

而
$$t=\frac{x}{\sqrt{y/n}}=\sqrt{n}\,\mathrm{tg}\,\theta,\ \cos\theta=\left(1+\frac{t^2}{n}\right)^{-\frac{1}{2}}$$

$$dt=\sqrt{n}\sec^2\theta\, d\theta=\sqrt{n}\left(1+\frac{t^2}{n}\right)d\theta$$

由此得到 T 的分布密度函数是

$$t(x;n)=c_1\left(1+\frac{x^2}{n}\right)^{-\frac{n-1}{2}}\left[\sqrt{n}\left(1+\frac{x^2}{n}\right)\right]^{-1}$$

$$=\frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{n}{2}\right)\sqrt{n}}\left(1+\frac{x^2}{n}\right)^{-\frac{n+1}{2}}$$

$$=\frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\sqrt{n\pi}}\left(1+\frac{x^2}{n}\right)^{-\frac{n+1}{2}}$$

图1给出了当 $n=1, 5, 10, \infty$ 时的 $t(n)$ 分布的密度函数图象.

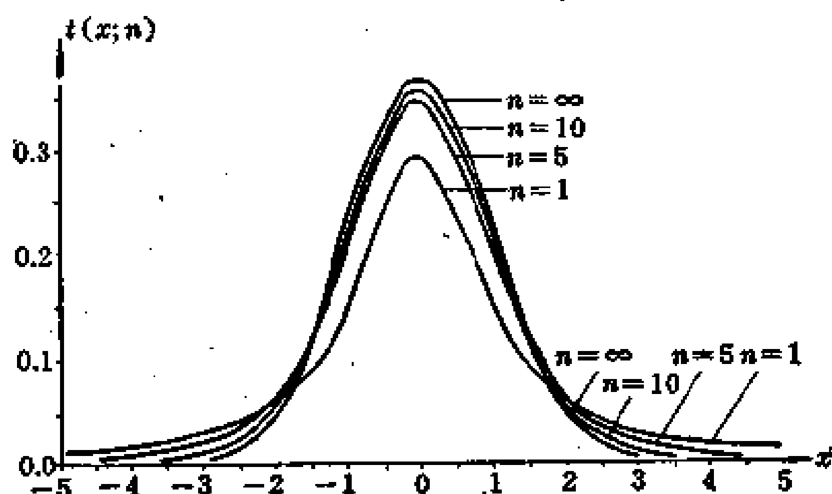


图1 t-分布密度曲线

系1 设 $X\sim N(\mu, \sigma^2)$, $Y/\sigma^2\sim\chi^2(n)$, 且 X 与 Y 相互独立, 则

$$T=\frac{X-\mu}{\sqrt{Y/n}}\sim t(n) \tag{6}$$

由(2)式可见, t-分布的密度函数 $t(x; n)$ 关于 x 是对称的, 且 $\lim\limits_{|x|\to\infty} t(x; n)=0$. 当 n 很大时, t-分布很接近于正态分布. 事实上,

$$\lim_{n\to\infty}\left(1+\frac{x^2}{n}\right)^{-(n+1)/2}=e^{-x^2/2}$$

然而对于比较小的 n 值, t-分布与正态分布之间存在有较大的差异. 而且有

$$P\{|T|\geqslant t_0\}\geqslant P\{|X|\geqslant t_0\}$$

其中 $X\sim N(0, 1)$. 也就是说, 在 t-分布的尾部比在标准正态分布的尾部有着更大的概率(见图2).

t-分布只存在阶数 $k<n$ 的矩, 例如当 $n=1$ 时, 就根本没有矩存在. 可以验证, 当 $n=1$ 时, t-分布是柯西分布的一个特例, 大家已经知道, 柯西分布是不存在任何阶矩的.

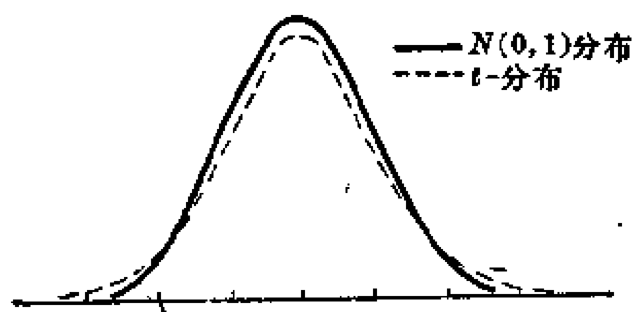


图 2

定理 2 设 $X \sim t(n)$, $n>1$, 则对 $r<n$, EX^r 存在, 且

$$EX^r=\begin{cases}0, & \text{当 } r \text{ 是奇数} \\ n^{r/2}\dfrac{\Gamma[(r+1)/2]\Gamma[(n-r)/2]}{\Gamma(1/2)\Gamma(n/2)}, & \text{当 } r \text{ 是偶数}\end{cases} \quad (7)$$

系 2 设 $X \sim t(n)$, 如果 $n>2$, 则 $EX=0$, $D(X)=EX^2=n/(n-2)$.

***二、非中心 t-分布**

定义 2 设 $X \sim N(\mu, \sigma^2)$, $Y/\sigma^2 \sim \chi^2(n)$, $\mu \neq 0$, 且 X 和 Y 独立. 则称

$$T=\frac{X}{\sqrt{Y/n}}$$

所服从的分布为非中心 t-分布, 带有非中心参数 $\delta=\mu/\sigma$ 和自由度 n.

非中心 t-分布的分布密度函数是

$$t(x; n, \delta)=\frac{n^{\frac{n}{2}}e^{-\frac{\delta^2}{2}}}{\sqrt{\pi}\,\Gamma\left(\frac{n}{2}\right)(n+x^2)^{\frac{n+1}{2}}}\sum_{m=0}^{\infty}\Gamma\left(\frac{n+m+1}{2}\right) \times\left(\frac{\delta^m}{m!}\right)\left(\frac{\sqrt{2}\,x}{\sqrt{n+x^2}}\right)^m \quad (8)$$

非中心 t 分布包含有 δ 和 n 两个参数, 当 $\delta=0$ 时, 由(8)式决定的 $t(x; n, 0)$ 就相同于(2)式的 $t(x; n)$ 分布, 所以 $t(x; n)$ 又称为中心 t-分布.

可以计算，如果 T 具有非中心 t-分布，其自由度为 n、非中心参数为 δ，则

$$ET=\delta\frac{\Gamma[(n-1)/2]}{\Gamma(n/2)}\sqrt{\frac{n}{2}},\quad n>1 \tag{9}$$

和

$$D(T)=\frac{n(1+\delta^2)}{n-2}-\frac{\delta^2 n}{2}\left(\frac{\Gamma[(n-1)/2]}{\Gamma(n/2)}\right)^2,\quad n>2 \tag{10}$$

三、F-分布

定义 3 设 X 和 Y 是相互独立的 χ^2-分布随机变量. 自由度分别为 m 和 n，则称随机变量

$$F=\frac{X/m}{Y/n}=\frac{X}{Y}\cdot\frac{n}{m} \tag{11}$$

所服从的分布为 F-分布，$(m,\ n)$ 称为它的自由度，且通常写为 $F\sim F(m,\ n)$.

定理 3 由(11)式所定义的 F 的分布密度函数是

$$f(x;m,\ n)=\begin{cases}\dfrac{\Gamma[(m+n)/2]}{\Gamma(m/2)\Gamma(n/2)}\left(\dfrac{m}{n}\right)\left(\dfrac{m}{n}x\right)^{\frac{m}{2}-1} & \\ \times\left(1+\dfrac{m}{n}x\right)^{-\frac{m+n}{2}}, & \text{当 } x>0\\ 0, & \text{当 } x<0\end{cases} \tag{12}$$

［证明］ 因为 $X\sim\chi^2(m)$，$Y\sim\chi^2(n)$，且 X 和 Y 独立，所以 $(X,\ Y)$ 的联合分布密度函数是

$$f(x,\ y)=\frac{1}{2^{\frac{m+n}{2}}\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)}e^{-\frac{x+y}{2}}x^{\frac{m}{2}-1}y^{\frac{n}{2}-1},\quad \text{对 } x>0,\ y>0$$

作变换

$$\begin{cases}u=x+y\\ v=\dfrac{x}{y}\cdot\dfrac{n}{m}\end{cases}$$

则

$$dxdy=\frac{m}{n}\frac{u}{\left(1+\frac{m}{n}v\right)^2}dudv$$

所以(U, V)的联合分布密度函数是

$$f_1(u, v)=\frac{1}{2^{\frac{m+n}{2}}\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)}e^{-\frac{u}{2}}\cdot u^{\frac{m+n}{2}-2}\left(\frac{m}{n}\right)^{\frac{m}{2}-1}$$

$$\times\frac{v^{\frac{m}{2}-1}}{\left(1+\frac{m}{n}v\right)^{\frac{m+n}{2}-2}}\cdot\frac{m}{n}\cdot\frac{u}{\left(1+\frac{m}{n}v\right)^2}$$

$$=\frac{1}{2^{\frac{m+n}{2}}\Gamma\left(\frac{m+n}{2}\right)}e^{-\frac{u}{2}}u^{\frac{m+n}{2}-1}\cdot\frac{\Gamma\left(\frac{m+n}{2}\right)\left(\frac{m}{n}\right)^{\frac{m}{2}}}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)}$$

$$\times\frac{v^{\frac{m}{2}-1}}{\left(1+\frac{m}{n}v\right)^{\frac{m+n}{2}}}$$

显然U和V相互独立，而且U服从$\chi^2(m+n)$分布，V的分布密度函数是

$$f(v)=\frac{\Gamma\left(\frac{m+n}{2}\right)\left(\frac{m}{n}\right)^{\frac{m}{2}}}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)}\frac{v^{\frac{m}{2}-1}}{\left(1+\frac{m}{n}v\right)^{\frac{m+n}{2}}}$$

因为$F=V$，所以F服从的分布是由分布密度函数(12)所决定的分布。

系3 设$X_1\sim\chi^2(m)$，$X_2\sim\chi^2(n)$，且X_1，X_2相互独立，则$Y=X_1+X_2$与$Z=X_1/X_2$相互独立。

注意一下定理3的证明过程，在这里$Y=U$，$Z=V\cdot\frac{m}{n}$，所以由U和V的独立性立即获得Y与Z相互独立。

系4 如果$X/\sigma^2\sim\chi^2(m)$，$Y/\sigma^2\sim\chi^2(n)$，且相互独立，则$F=\frac{X}{Y}\cdot\frac{n}{m}\sim F(m, n)$分布。

系5 如果$X\sim F(m, n)$分布，则$1/X\sim F(n, m)$分布。

如果我们在(11)式中取 X 是自由度为 1 的 χ^2 分布，即 $m=1$，则 $F=[t(n)]^2$. 所以 $F(1, n)$ 与 $t^2(n)$ 有相同的分布密度函数.

图 3 给出了 F 分布的密度函数.

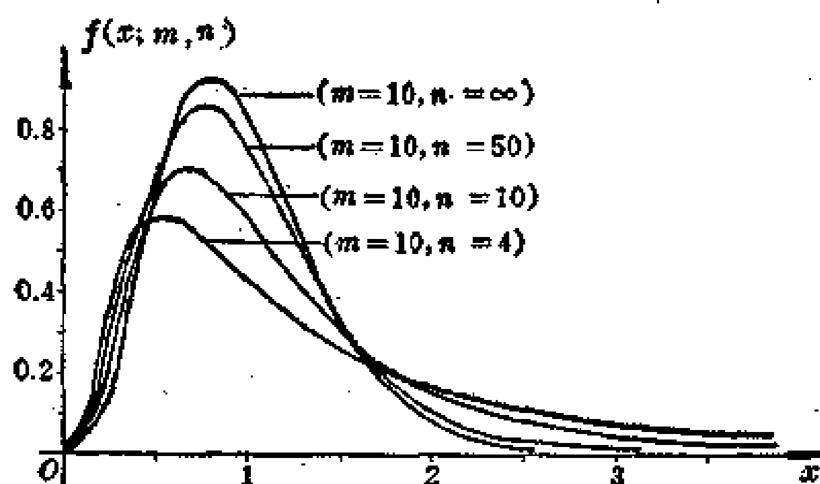


图 3　F-分布密度函数曲线

定理 4　设 $X\sim F(m, n)$，则对 $r>0$ 有

$$EX^r=\left(\frac{n}{m}\right)^r\frac{\Gamma\left(r+\frac{m}{2}\right)\Gamma\left(\frac{n}{2}-r\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)},\quad \text{对 } 2r<n \tag{13}$$

特别有

$$EX=\frac{n}{n-2},\quad \text{对 } n>2 \tag{14}$$

和

$$D(X)=\frac{n^2(2m+2n-4)}{m(n-2)^2(n-4)},\quad \text{对 } n>4 \tag{15}$$

证明留给读者作练习.

定理 5　设 $X_1, \cdots, X_n$ 为相互独立，同服从于 $N(0, \sigma^2)$ 的随机变量.

又设 Q_i, $i=1, \cdots, k$ 是秩为 n_i 的非负二次型. 且

$$Q_1+\cdots+Q_k=\sum_{i=1}^{n}X_i^2$$

$$n_1+\cdots+n_k=n$$

则
$$F_{ij}=\frac{Q_i}{Q_j}\cdot\frac{n_j}{n_i}$$
服从自由度为(n_i, n_j)的 F-分布.

由定理 2.4 和 F 分布的定义立即可得到定理 5.

定理 5 与定理 2.4 一起在 § 4.3 方差分析中发挥重要作用.

*四、非中心 F-分布

定义 4 设 $X\sim\chi^2(m, \delta)$, $Y\sim\chi^2(n)$, $\delta\neq0$, 且设 X 和 Y 相互独立. 则称
$$F=\frac{X/m}{Y/n} \tag{16}$$
所服从的分布为非中心 F-分布, 带有自由度(m, n)和非中心参数 δ.

由(16)所定义的 F 变量的分布密度函数是

$$f(x; m, n, \delta)=\begin{cases}\dfrac{m^{\frac{m}{2}}n^{\frac{n}{2}}}{\Gamma\left(\frac{n}{2}\right)}e^{-\frac{\delta}{2}}x^{\frac{m}{2}-1}\sum\limits_{k=0}^{\infty}\dfrac{\left(\frac{\delta mx}{2}\right)^k\Gamma\left(\frac{m+n}{2}+k\right)}{k!\Gamma\left(\frac{m}{2}+k\right)(mx+n)^{\frac{m+n}{2}+k}}, & \text{当 } x>0\\ 0, & \text{当 } x<0\end{cases} \tag{17}$$

它含有 δ 和 n 两个参数, 如果 $\delta=0$, (17)就简化为 F-分布(12), 故 F-分布又称为中心 F-分布.

非中心 F-分布的数学期望和方差是
$$E(F)=\frac{n(m+\delta)}{m(n-2)}, \quad \text{对 } n>2$$
和
$$D(F)=\frac{2n^2}{m^2(n-2)^2(n-4)}[(m+\delta)^2+(n-2)(m+2\delta)],$$
$$\text{对 } n>4$$

§ 4. 正态母体子样均值和方差的分布

[例 1] 设母体 X 服从正态分布 $N(0, \sigma^2)$. 从该母体中抽

取一个容量为 n 的简单子样，它的观察值为 $(x_1, x_2, \cdots, x_n)$，从这些子样值按公式计算

$$\bar{x}=\frac{1}{n}\sum_{k=1}^{n} x_k$$

$$s_n^2=\frac{1}{n}\sum_{k=1}^{n}(x_k-\bar{x})^2=\frac{1}{n}\sum_{k=1}^{n}x_k^2-\bar{x}^2$$

由此算得的 $(\bar{x}, s_n^2)$ 是二维随机变量 $(\bar{X}, S_n^2)$ 的观察值，这儿

$$\bar{X}=\frac{1}{n}\sum_{k=1}^{n}X_k,\ S_n^2=\frac{1}{n}\sum_{k=1}^{n}(X_k-\bar{X})^2$$

在实际应用中经常遇到下面的问题：在我们通过子样观察值算得 $(\bar{X}, S_n^2)$ 的观察值 $(\bar{x}, s_n^2)$ 后，需要知道使得不等式 $\bar{X}>\bar{x}$ 和 $S_n^2\geqslant s_n^2$ 同时成立这一事件的概率. 为此就必须找出统计量 $(\bar{X}, S_n)$ 的联合分布.

定理 1 设 $X_1, X_2, \cdots, X_n$ 是从正态母体 $N(\mu, \sigma^2)$ 中抽取的一个简单子样. 记

$$\bar{X}=\frac{1}{n}\sum_{k=1}^{n}X_k,\ S_n^2=\frac{1}{n}\sum_{k=1}^{n}(X_k-\bar{X})^2 \tag{1}$$

则有 (1) $\bar{X}$ 和 S_n^2 独立；

(2) $\bar{X}\sim N\left(\mu, \frac{1}{n}\sigma^2\right)$； (2)

(3) $nS_n^2/\sigma^2\sim\chi^2(n-1)$. (3)

［证明］ 记 A 是一 $n\times n$ 正交矩阵，使得它的第一行是 $(n^{-\frac{1}{2}}, n^{-\frac{1}{2}}, \cdots, n^{-\frac{1}{2}})$，作正交变换

$$\begin{pmatrix}Y_1\\Y_2\\\vdots\\Y_n\end{pmatrix}\triangleq A\begin{pmatrix}X_1\\X_2\\\vdots\\X_n\end{pmatrix}$$

则有 (i)
$$Y_1=\frac{1}{\sqrt{n}}\sum_{k=1}^{n}X_k=\sqrt{n}\,\bar{X} \tag{4}$$

(ii) $$Y_1^2+Y_2^2+\cdots+Y_n^2=X_1^2+X_2^2+\cdots+X_n^2$$
$$=\sum_{k=1}^{n}(X_k-\overline{X})^2+n\overline{X}^2$$

即 $$Y_2^2+Y_3^2+\cdots+Y_n^2=\sum_{k=1}^{n}(X_k-\overline{X})^2=nS_n^2 \tag{5}$$

(iii) $$(X_1-\mu)^2+\cdots+(X_n-\mu)^2$$
$$=(Y_1-\sqrt{n}\,\mu)^2+Y_2^2+\cdots+Y_n^2 \tag{6}$$

与定理1.2类似地可证明 $Y_1, \cdots, Y_n$ 也是独立正态随机变量，且

$$D(Y_k)=\sigma^2,\ k=1,\ \cdots,\ n \tag{7}$$

$$EY_1=\sqrt{n}\,\mu \tag{8}$$

$$EY_k=0,\ k=2,\ 3,\ \cdots,\ n \tag{9}$$

所以 Y_1 与 nS_n^2 独立；

$$Y_1\sim N(\sqrt{n}\,\mu,\ \sigma^2)$$

$$nS_n^2/\sigma^2\sim\chi^2(n-1)$$

而 $$\overline{X}=\frac{1}{\sqrt{n}}Y_1$$

故 $\overline{X}$ 与 S_n^2 独立； $\overline{X}\sim N\left(\mu,\ \frac{1}{n}\sigma^2\right)$; $nS_n^2/\sigma^2\sim\chi^2(n-1)$.

系1 $$T=\frac{\overline{X}-\mu}{S_n}\sqrt{n-1}\sim t(n-1)$$

[证明] 因为 $\frac{\sqrt{n}(\overline{X}-\mu)}{\sigma}\sim N(0,\ 1)$, $nS_n^2/\sigma^2\sim\chi^2(n-1)$,

所以

$$T=\frac{\dfrac{\sqrt{n}\,(\overline{X}-\mu)}{\sigma}}{\sqrt{\dfrac{nS_n^2}{\sigma^2(n-1)}}}=\frac{(\overline{X}-\mu)}{S_n}\sqrt{n-1} \tag{10}$$

服从自由度为 $n-1$ 的 t–分布．

系2 设 $X_1,\ X_2,\ \cdots,\ X_m$ 是从正态母体 $N(\mu_1,\ \sigma_1^2)$抽取的一个子样；$Y_1, Y_2,\ \cdots,\ Y_n$ 是从正态母体 $N(\mu_2,\ \sigma_2^2)$ 抽取的一个

子样，且假定 $X_1, \cdots, X_m$ 和 $Y_1, \cdots, Y_n$ 相互独立. 则

$$F=\frac{mS_{1m}^2}{nS_{2n}^2}\cdot\frac{(n-1)\sigma_2^2}{(m-1)\sigma_1^2} \tag{11}$$

服从 $F(m-1, n-1)$ 分布. 其中

$$\begin{aligned} S_{1m}^2&=\frac{1}{m}\sum_{k=1}^{m}(X_k-\bar{X})^2, \quad \bar{X}=\frac{1}{m}\sum_{k=1}^{m}X_k \\ S_{2n}^2&=\frac{1}{n}\sum_{k=1}^{n}(Y_k-\bar{Y})^2, \quad \bar{Y}=\frac{1}{n}\sum_{k=1}^{n}Y_k \end{aligned} \tag{12}$$

特别地，如果 $\sigma_1^2=\sigma_2^2$，则 $F=\dfrac{m(n-1)S_{1m}^2}{n(m-1)S_{2n}^2}\sim F(m-1, n-1)$.

系 3 设 $X_1, \cdots, X_m$ 和 $Y_1, \cdots, Y_n$ 分别是从正态母体 $N(\mu_1, \sigma^2)$ 和 $N(\mu_2, \sigma^2)$ 中所抽取的独立子样. 则

$$T=\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{mS_{1m}^2+nS_{2n}^2}}\cdot\sqrt{\frac{mn(m+n-2)}{m+n}} \tag{13}$$

服从于 $t(m+n-2)$ 分布. 此结论的证明留给读者作练习.

这里需注意，它要求两个正态母体分布的方差相等.

习　题

1. 设 $X_1, \cdots, X_n$ 是 X 的 n 个独立观察，试求子样均值 $\bar{X}$ 的概率分布或分布密度，如果(1) $X\sim P(\lambda)$; (2) $X\sim\chi^2(m)$.

2. 设 $X_1, \cdots, X_n$ 是从服从贝努利分布 $b(1, p)$ 中抽取的简单随机子样，$0<p<1$. $\bar{X}$ 是子样均值，S_n^2 是子样方差，试求 S_n^2 的概率分布. (提示：$S_n^2=\bar{X}(1-\bar{X})$)

3. Γ-分布 $G(\alpha, p)$. ($\alpha>0, p>0$) 如果随机变量 X 的分布密度函数是

$$g(x;\alpha, p)=\begin{cases}\dfrac{\alpha^p}{\Gamma(p)}e^{-\alpha x}x^{p-1}, & x>0\\ 0, & x<0\end{cases}$$

则称 X 服从 Γ-分布 $G(\alpha, p)$. 求

(1) $G(\alpha, p)$ 的特征函数是 $\varphi(t)=\left(\dfrac{\alpha}{\alpha-it}\right)^p$;

(2) $E(X)=\dfrac{p}{\alpha}$, $D(X)=\dfrac{p}{\alpha^2}$;

(3) 如果 $X_i \sim G(a, p_i)$, $i=1, 2, \cdots, k$, 且 $X_1, \cdots, X_k$ 相互独立, 记 $p=\sum_{i=1}^{k} p_i$. 则 $\sum_{i=1}^{k} X_i \sim G(a, p)$.

(4) 如果取 $a=1/2$, $p=\frac{n}{2}$, 则 $G\left(\frac{1}{2}, \frac{n}{2}\right)$ 是 $\chi^2(n)$ 分布.

4. 如果 X 服从非中心 χ^2-分布 $\chi^2(n, \delta)$, 计算

(1) X 的特征函数 $\varphi(t)=(1-2it)^{-n/2}\exp\left\{\frac{it\delta}{1-2it}\right\}$;

(2) $EX=n+\delta$, $D(X)=2n+4\delta$;

(3) 证明 如果 $X_i \sim \chi^2(n_i, \delta_i)$, 且 $X_1, \cdots, X_k$ 相互独立, 则

$$\sum_{i=1}^{k} X_i \sim \chi^2(n, \delta)$$

其中

$$n \triangleq \sum_{i=1}^{k} n_i, \quad \delta \triangleq \sum_{i=1}^{k} \delta_i$$

5. 设 $X^\tau=(X_1, \cdots, X_n)$ 服从 n 元正态分布, $EX=\mu$, $\mathrm{cov}(X, X)=\Sigma$, 且 Σ^{-1} 存在, 则 $U=(X-\mu)^\tau \Sigma^{-1}(X-\mu)$ 服从于自由度为 n 的 χ^2-分布.

6. 说明非中心 t-统计量的平方是非中心 F-统计量.

7. 设 $X_1, \cdots, X_n$ 是从正态 $N(\mu, \sigma^2)$ 分布的母体中抽取的简单子样, $\overline{X}$ 和 S_n^2 分别表示它的子样均值和子样方差. 又设 $X_{n+1} \sim N(\mu, \sigma^2)$, 且与 $X_1, \cdots, X_n$ 独立. 试求统计量

$$\frac{X_{n+1}-\overline{X}}{S_n}\sqrt{\frac{n-1}{n+1}}$$

的抽样分布.

8. 设 $X_1, \cdots, X_m$ 和 $Y_1, \cdots, Y_n$ 分别是从 $N(\mu_1, \sigma^2)$ 和 $N(\mu_2, \sigma^2)$ 的母体中抽取的独立随机子样, $\overline{X}$ 和 $\overline{Y}$ 分别表示 X 和 Y 的子样均值, S_{1m}^2 和 S_{2n}^2 分别表示 X 和 Y 的子样方差. α 和 β 是两个固定的实数, 试求

$$\frac{\alpha(\overline{X}-\mu_1)+\beta(\overline{Y}-\mu_2)}{\sqrt{\frac{mS_{1m}^2+nS_{2n}^2}{m+n-2}}\sqrt{\frac{\alpha^2}{m}+\frac{\beta^2}{n}}}$$

的抽样分布.

第三章 假设检验(I)

§1. 引 言

在第一章中介绍了两种常用的参数估计方法，实践中还提出另一类很重要的统计推断问题，它就是本章要讨论的假设检验问题．先考察几个例子．

[例 1] 某厂有一批产品，共一万件，须经检验后方可出厂．按规定标准，次品率不得超过5%．今在其中任意选取50件产品进行检查，发现有次品4件，问这批产品能否出厂．

在这个例子中，我们事先对这批产品次品率的情况一无所知．当然，从频率稳定性来说，我们可以用被检查的50件产品的次品率4/50来估计这整批产品的次品率，但是我们目前所关心的问题是：如何根据抽样的次品率$\nu/n(=4/50)$推断整批产品的次品率是否超过了5%．也就是说，首先我们可以对整批产品作一种假设：次品率低于5%，然后利用子样中的次品率ν/n来检验我们所作这一假设的正确性．

[例 2] 糖厂用自动包装机将糖装箱，以利外运．每箱的标准重量规定为100斤．每天开工时，需要先检验一下包装机工作是否正常．根据以往的经验知道，用自动包装机装箱，其各箱重量的标准差$\sigma=1.15$斤．某日开工后，抽测了九箱，其重量如下(单位：斤)：

99.3, 98.7, 100.5, 101.2, 98.3, 99.7, 99.5, 102.1, 100.5

试问此包装机工作是否正常．

在这个例子中，我们关心的问题是：包装机工作是否正常，即

包装机装出的糖箱的平均重量是否符合标准为 100 斤. 一般认为自动包装机装箱其重量的起伏是服从正态分布的. 因此此例可作如下处理: 先假设母体(包装机装出的糖箱)的平均值 $\mu=100$(斤), 然后利用上述抽取的九个数据, 来推断我们所作这一假设的正确性,从而拒绝或接受这种假设.

[例 3] 某种建筑材料, 其抗断强度的分布以往一直符合正态分布,今改变了配料方案,希望确定其抗断强度的分布是否仍为正态分布?

与前 2 例类似,先建立假设:假设改变配料方案后生产出的该建筑材料的抗断强度仍服从正态分布. 然后通过抽取子样来推断上述这种假设的正确性.

上述三例的共同特点是: 先对母体的分布函数的形式或分布函数的某些参数作某种假设, 然后抽取子样和集中子样中的有关信息,对假设的正确性进行推断. 今后,我们把任何一个在母体的未知分布上所作的假设称为统计假设. 并记为 H_0. 对上面所举的三个例子,统计假设分别是: H_0: p(次品率)$\leqslant 0.05$; H_0: μ(均值)$=100$(斤); H_0: $F(x)$(分布函数)$\in N(\cdot,\cdot)$(它表示 $F(x)$ 属于正态分布函数族).

这里可以看到,例 1 与例 2 均给出了母体分布的形式,统计假设 H_0 是对未知参数作的.而例 3 与例1、2不同, 它的统计假设 H_0 是直接对母体的分布函数给出的. 产生这一差别的主要原因是由于我们事先对于母体分布函数的知识具有很大差别. 在例 1 和例 2 里对母体的真分布具有较多的知识, 已知它的函数形式, 未知的仅是一个或几个参数,一旦这些未知参数知道后,母体的真分布就完全已知了. 例如在例 1 中如果知道了次品率 p 的具体数值为 p_0, 那末母体的真分布就完全确定了, 是贝努里分布 $b(1, p_0)$; 在例 2 中如果知道母体的均值 $\mu=100$(斤), 那么就知道母体的真分

布是 $N(100, 1.15^2)$. 正由于母体的真分布完全被几个未知参数所决定. 因此任何一个关于母体未知分布的假设总可以等价地给出在它的未知参数上. 这种仅涉及到母体分布中所包含的几个未知参数的统计假设称为**参数假设**.

但对有些实际问题,我们事先很少知道关于母体分布的知识,如在例3中只知道改变配方后的材料的抗断强度服从于连续型分布,而不知道它的分布所具有的函数形式. 因此统计假设只能直接给出在未知分布函数的形式上或是它的某些数字特征上,这种统计假设显然不同于前述的参数假设,我们称这种假设为**非参数假设**. 在例3中,H_0: $F(x) \in N(\cdot, \cdot)$,就是一种非参数假设.

对于一个假设检验问题,首先是根据实际问题的要求提出统计假设 H_0,但这仅是第一步,提出统计假设的目的是要求进一步推断所提出的统计假设 H_0 是否正确. 这就要求建立推断统计假设 H_0 的方法. 在统计学上,称判断给定统计假设 H_0 的方法为**统计假设检验**,或简称为**统计检验**.

另外,在有些实际问题中,只提出一个统计假设,如例1～例3那样. 但是在另一些实际问题中,往往需提出两个甚至多个统计假设,而且统计检验的目的也需要同时判断多个假设中哪一个是正确的. 如果一个统计问题中仅提出一个统计假设,而且我们的目的也仅仅是判断这一个统计假设是否成立,并不同时研究其它统计假设. 这类检验问题称为**显著性检验**. 本章将集中地讨论显著性检验方法,特别是正态母体的参数的显著性检验问题.

怎样检验一个统计假设呢? 在开展一般讨论之前先分析一个具体例子,并引入一些概念.

[例4] 设母体 X 服从正态分布 $N(\mu, 1)$,其中仅包含一个未知参数,即数学期望 μ,欲要求检验统计假设 H_0: $\mu=0$.

在这里,母体 X 的分布的函数形式是已知的,为正态分布

$N(\mu, 1)$，其中仅含一个未知参数 μ. 同时也仅提出一个统计假设 $H_0: \mu=0$. 所以它是一个参数的显著性检验问题.

怎样判断这一统计假设 $H_0(\mu=0)$ 的正确性呢？首先需对母体进行一定次数的观察，获得数据，也就是说抽取子样. 设我们从该母体中抽取了一个容量为10的简单随机子样，子样观察值记为 $(x_1, \cdots, x_{10})$. 子样来自母体，反映了母体的分布规律，因此子样中必然包含关于未知参数 μ 的信息. 但是要从子样直接推断统计假设是否成立是困难的，还必需对子样进行加工，把子样中包含的关于未知参数 μ 的信息集中起来，也就是说要构造一个适用于检验假设 H_0 的统计量.

这里 μ 是母体的均值，第一章中已经看到，子样均值 $\bar{X}$ 是 μ 的一个无偏估计，且其方差

$$D(\bar{X})=\frac{1}{n}D(X)=\frac{1}{n}=\frac{1}{10}$$

这说明子样均值 $\bar{X}$ 比子样的每个分量 X_i 更集中地分布在母体均值 μ 的周围，如果假设 $H_0(\mu=0)$ 是真的，则子样均值 $\bar{X}$ 的观察值应较集中在0点附近，否则就应有偏离0点的趋势. 这表明子样均值 $\bar{X}$ 较好地集中了子样中所包含的关于 μ 的信息. 因此利用 $\bar{X}$ 构造判断统计假设 $H_0(\mu=0)$ 的方法是合适的.

若从子样观察值计算得到的子样均值 $\bar{X}$ 的观察值 $\bar{x}=1.01$，那对假设 $H_0(\mu=0)$ 的正确性能作出怎样的判断呢？

当假设 $H_0(\mu=0)$ 成立时，X 服从于 $N(0, 1)$ 分布. 由系2.1.1知道，$\bar{X}$ 服从 $N\left(0, \frac{1}{n}\right)$ 分布，因而 $\bar{X}$ 的绝对值大于等于上述观察值 $\bar{x}=1.01$ 的概率是

$$\begin{aligned} P\{|\bar{X}|\geqslant 1.01\} &= P\{|\bar{X}|\sqrt{10}\geqslant 1.01\sqrt{10}\} \\ &\approx 2[1-\Phi(3.05)]\approx 2(1-0.999)=0.002 \end{aligned}$$

这表明如果假设 H_0 成立，那么事件 $\{|\bar{X}|\geqslant 1.01\}$ 实际上不大可

能出现，因为这事件的概率大约为 0.002，即在 1000 次中，大约只有 2 次使我们所观察到的子样平均值大于等于 1.01.

现在是否可以根据事件$\{\overline{X}\geqslant 1.01\}$的概率很小这一理由而证实 H_0 不成立呢？当然不能就此而作出 H_0 不成立这一肯定的结论．因为在假设 H_0 成立的条件下，事件$\{|\overline{X}|\geqslant 1.01\}$的概率虽很小，但这个事件仍可能出现，只是可能性很小而已．显然我们也不能作肯定 H_0 成立的结论．但是现在必须根据这一子样观察作出一个判断，即必须决定是接受还是拒绝假设 $H_0(\mu=0)$．所以只能从"拒绝"和"接受"中选择一个较为合理的判断作为我们的决定．一般它可以这样处理：给定一个临界概率 α，如果在假设 H_0 成立的条件下，出现观察到的事件 $\{|\overline{X}|\geqslant \bar{x}\}$的概率小于等于 α，就作拒绝假设 H_0 的决定.

这个临界概率 α 在假设检验里是重要的，我们把这个 α 称为显著水平或简称为水平，$0<\alpha<1$．对于各种不同的问题，显著水平 α 可以选取得不一样，但一般应取 α 为一个较小的数．这是因为我们给出假设 H_0 是经过细致的调查和考察的．所以对假设 H_0 需加以保护，也就是说拒绝它应该慎重．为查表方便通常我们取 α 为一些标准值，如 $\alpha=0.05$ 或 0.01 等．这表明当假设 H_0 为真时，出现这事件的可能性大约在 100 次中不超过 5 次（对 $\alpha=0.05$）或 1 次（对 $\alpha=0.01$），这是一个小概率事件．根据小概率事件在一次试验（观察）中不可能发生的实际推断原理，如果出现了这事件我们就有理由怀疑 H_0 不真．因为它超出了在 H_0 成立条件下能以随机波动来解释的范围，因而作出拒绝 H_0 的判断.

尚须指出，虽然在假设 H_0 为真时，发生作出拒绝 H_0 这一错误判断的概率很小，它小于等于 α．但这一错误还是可以发生的。在统计学上，称这种类型的错误为第一类错误．所以显著水平 α 是用来控制犯第一类错误的．显然，如果 α 取得大，发生拒绝 H_0

的可能性就大，也就是说需对较多的子样观察值作出拒绝 H_0 的判断.

在例 4 中，如果取 $\alpha=0.01$，我们应作拒绝 $H_0(\mu=0)$ 的判断. 当然取 $\alpha=0.05$ 时，更应拒绝 H_0.

[例 5] 设从带有未知均值 μ 的一个正态 $N(\mu, 1)$ 母体中抽取一个容量 $n=16$ 的随机子样，算得子样均值的观察值 $\bar{x}=0.1$，要求判断统计假设 $H_0: \mu=0$.

如果假设 $H_0(\mu=0)$ 成立，则 $\overline{X}$ 服从 $N(0, 1/16)$ 分布，查附表 1 得到

$$P\{|\overline{X}|\geqslant 0.1\}=P\{4|\overline{X}|\geqslant 0.4\}=2[1-\Phi(0.4)]$$
$$\approx 2\times 0.345=0.690$$

在这里，$|\overline{X}|$ 大于等于观察值的概率比 $\alpha=0.05$ 大得多；根据上面所述的处理方法，现在我们不能拒绝假设 $H_0(\mu=0)$，但并不能由此不另加分析地得出假设 H_0 是正确的. 一个显著性检验一般只能作出一方面的决定，如果在假设 H_0 为正确的假定下，出现观察到的事件的概率小于或等于 α，那么我们能够拒绝 H_0；如果出现观察到的事件的概率大于 α，那么我们只能说，实验结果与假设无显著差异. 从概率论的角度讲，根据在假设 H_0 成立的条件下概率大于 0.05(或者甚至大于 0.01)的事件在一次试验中出现就决定接受假设 H_0 还是有些武断的. 我们特别提出这一点，因为往往发生这样的误解：经过一个显著性检验没有被拒绝的假设误解为一定是正确的假设.

从这里看到，在统计假设检验里，通过子样观察，我们必需对被检验的假设作出明确的判断，是接受还是拒绝这假设. 但是我们也必需承认这种判断有可能发生错误. 至于发生错误的概率有多大，错误的概率与显著水平 α 之间有什么关系等等问题将在第六章中系统讨论.

由上讨论可见显著性检验问题的处理步骤是：

(1) 建立统计假设 H_0；

(2) 构造一个合适的统计量 U 和从子样观察值计算出统计量 U 的观察值 u；

(3) 规定一个显著水平 α(一般取 0.05 或 0.01)，求出在 H_0 成立的条件下能使 $P_{H_0}\{|U|\geqslant u_0\}\leqslant\alpha$ 满足的值 u_0；

(4) 比较观察值 u 和 u_0，如果 $|u|\geqslant u_0$，则拒绝假设 H_0.

上面看到，寻找检验统计量 U 的分布，至少对于给定的 α 要找出满足 $P_{H_0}\{|U|\geqslant u_0\}=\alpha$ 的临界值 u_0 是很重要的．按进行检验时所取的子样容量的大、小，分为小样和大样两类问题，对于小样的显著性检验，需要给出检验统计量 U 的精确分布，而对于大样问题可利用 U 的极限分布作为近似.

§2. 正态母体参数的检验

现在利用第二章中所讨论的抽样分布，详细地介绍几种常用的检验法.

一、均值的检验

设 $X_1, X_2, \cdots, X_n$ 是从正态 $N(\mu, \sigma_0^2)$ 母体中抽取的一个子样，其中 σ_0^2 是已知常数，今欲检验假设 $H_0: \mu=\mu_0$.

由 §1 的讨论知道，检验问题的关键是基于子样寻找一个合适的统计量．类似于例 1.4 的讨论，在这里子样平均值 $\overline{X}$ 很好地集中了子样中所包含的关于 μ 的信息．当假设 H_0 成立时 $\overline{X}$ 的观察值 $\bar{x}$ 较集中地分布在 μ_0 的周围，否则就有偏离 μ_0 的趋势．所以 $\overline{X}$ 可以用来检验假设 $H_0(\mu=\mu_0)$．为了查表方便将 $\overline{X}$ 标准化，得到统计量

$$U=\frac{\overline{X}-\mu_0}{\sigma_0}\sqrt{n} \tag{1}$$

在 $H_0(\mu=\mu_0)$ 为真时，U 服从 $N(0, 1)$ 分布，而当 H_0 不真时，U 服从均值不为零的 $N(\cdot,1)$ 分布，这表明当 H_0 不真时，$|U|$ 的观察值有偏大的趋势，所以对给定的显著水平 α，查正态分布表（见附表 1）求取 $u_{\alpha/2}$，使得

$$P_{H_0}\{|U|\geqslant u_{\alpha/2}\}=\sqrt{\frac{2}{\pi}}\int_{u_{\alpha/2}}^{\infty} e^{-t^2/2}dt=\alpha \qquad (2)$$

然后将子样观察值 $(x_1, \cdots, x_n)$ 代入(1)式算出 U 的观察值 u，并比较 u 和 $u_{\alpha/2}$，若 $|u|\geqslant u_{\alpha/2}$ 则拒绝假设 $H_0(\mu=\mu_0)$，即认为母体的真实均值 μ 与 H_0 给定的均值 μ_0 之间有显著差异，否则接受假设 H_0，即认为观察结果与假设 $H_0(\mu=\mu_0)$ 无显著差异．这个检验法称为 u-检验法．

［例 1］（续例 1．2）对糖厂的自动包装机，我们认为其装箱的重量的波动服从正态分布．例 1.2 又告诉我们根方差 $\sigma_0=1.15$（斤）是已知常数．此时自动包装机生产是否正常的标志是装出糖箱的重量的平均值是否为 100 斤，因此它归结为要求检验假设 $H_0: \mu=100$．

这里 $n=9$，$\sigma_0=1.15$．

$$\bar{x}=\frac{1}{9}(99.3+98.7+100.5+101.2+98.3$$

$$+99.7+99.5+102.1+100.5)$$

$$=\frac{1}{9}(899.8)=99.98$$

代入(1)式得 U 的观察值

$$u=\frac{\bar{x}-\mu_0}{\sigma_0}\sqrt{n}=\frac{(99.98-100)}{1.15}\sqrt{9}\approx-0.052$$

若给定 $\alpha=0.05$，查附表 1 得 $u_{\alpha/2}=1.96$，这时因

$$|u|=0.052<1.96=u_{\alpha/2}$$

所以不能否认 H_0 为真，故可认为此时包装机的工作正常．

u-检验法不仅用来研究一个正态母体，在方差已知的条件下其均值是否与已知常数相符合，而且也用来比较两个正态母体的均值是否相等．设两个正态分布的母体 $\mathbb{X}$ 和 $\mathbb{Y}$，在已知 $D(\mathbb{X})=\sigma_1^2$，$D(\mathbb{Y})=\sigma_2^2$ 的条件下，要求检验假设 H_0: $E\mathbb{X}=E\mathbb{Y}$，即比较两个正态母体的均值是否有显著差异．对这类问题的检验方法如下：

分别从两个正态母体抽取容量为 m 和 n 的子样，计算相应的子样平均值 $\overline{X}$ 和 $\overline{Y}$．考察 $\overline{X}-\overline{Y}$，有

$$D(\overline{X}-\overline{Y})=D(\overline{X})+D(\overline{Y})=\frac{1}{m}\sigma_1^2+\frac{1}{n}\sigma_2^2$$

关于它的均值，当 $H_0(E(\mathbb{X})=E(\mathbb{Y}))$ 成立时，有 $E(\overline{X}-\overline{Y})=0$，否则 $E(\overline{X}-\overline{Y})\neq 0$．将 $\overline{X}-\overline{Y}$ 标准化得到

$$U=\frac{\overline{X}-\overline{Y}}{\sqrt{\frac{1}{m}\sigma_1^2+\frac{1}{n}\sigma_2^2}} \tag{3}$$

当 $H_0(E(X)=E(Y))$ 成立时，U 服从 $N(0,1)$ 分布，否则它服从均值不为零的 $N(\cdot,1)$ 分布．所以与前述检验 $H_0(\mu=0)$ 一样，对给定的 α，由(2)式定出 $u_{\alpha/2}$，并把由子样观察值所计算得到的统计量(3)的观察值 u 与 $u_{\alpha/2}$ 比较，若 $|u|\geqslant u_{\alpha/2}$ 则拒绝 H_0．

u-检验法还可以应用于非正态母体的大子样检验问题：

［例 2］（续例 1.1）对上节的例 1，为简单起见，我们可将此问题归结为希望利用次品率 ν/n 来检验母体次品率 p 是否满足假设 H_0: $p=p_0(=0.05)$．

用 $\mathbb{X}$ 记母体元素的指标，有

$$\mathbb{X}=\begin{cases}0, & \text{好品}\\ 1, & \text{次品}\end{cases}$$

则在假设 H_0 成立时 $P\{\mathbb{X}=0\}=1-p_0$，$P\{\mathbb{X}=1\}=p_0$，$E\mathbb{X}=p_0$，$D(\mathbb{X})=p_0(1-p_0)$，设 $X_1, \cdots, X_n$ 是一子样，则

$$\overline{X}=\frac{1}{n}\sum_{i=1}^{n} X_i=\frac{\nu}{n}$$

其中 ν 表示子样中的次品数.

由中心极限定理知道，在 $H_0(p=p_0)$ 成立的条件下，

$$U=\frac{(\overline{X}-EX)}{\sqrt{D(X)}}\sqrt{n}=\frac{\left(\frac{\nu}{n}-p_0\right)}{\sqrt{p_0(1-p_0)}}\sqrt{n}$$

$$=\frac{(\nu-np_0)}{\sqrt{np_0(1-p_0)}} \tag{4}$$

渐近于 $N(0,1)$ 分布，因此当 n 较大时(一般在 30 以上)，可把(4)式决定的 U 近似地作为正态变量来处理.

现在 $p_0=0.05$, $n=50$, $\nu=4$, 代入(4)式得

$$u=\frac{\left(\frac{4}{50}-0.05\right)}{\sqrt{0.05\times0.95}}\sqrt{50}\approx0.96$$

对$\alpha=0.01$，查表得 $u_{\alpha/2}=2.58$，这时因

$$|u|=0.96<2.58=u_{\alpha/2}$$

所以不能拒绝假设 $H_0(p=0.05)$.

上面讨论的 u-检验法都是构造了一个统计量 U，它具有标准正态 $N(0,1)$ 分布(或者是精确分布或者是渐近分布)，为了标准化我们必需知道母体的方差. 但是在许多实际问题里，方差往往是未知的，那么我们怎样检验关于正态母体均值的有关假设呢？这就是下面的 t-检验法所要解决的问题.

设母体服从正态分布 $N(\mu, \sigma^2)$, μ, σ^2 都是未知参数, X_1, $\cdots$, X_n 是它的一个子样，欲检验假设 H_0: $\mu=\mu_0$.

现在由于母体方差 σ^2 是未知常数，因此(1)式不能应用(它不再是统计量). 一个很自然的想法是以子样方差来代替(1)式中的母体方差 σ^2，构造新的 t-统计量

$$T=\frac{\overline{X}-\mu_0}{S_n}\sqrt{n-1} \tag{5}$$

其中
$$S_n^2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2$$

由系 2.4.1 知道，在 H_0 为真时，$T\sim t(n-1)$，否则它服从非中心 t-分布．若给定显著水平 α，查 t-分布表可得 $t_{\alpha/2}$，使得

$$P\{|T|\geqslant t_{\alpha/2}\}=\alpha \tag{6}$$

(见图 1)．在附表 3 中，对某些特定选取的 α，给出了 $t_{\alpha/2}$ 的值．

由子样观察值，代入(1)式可算得 T 的观察值 t．若 $|t|\geqslant t_{\alpha/2}$，则拒绝假设 $H_0(\mu=\mu_0)$．因为这个检验统计量 T 具有 t-分布所以称为 t-检验法．

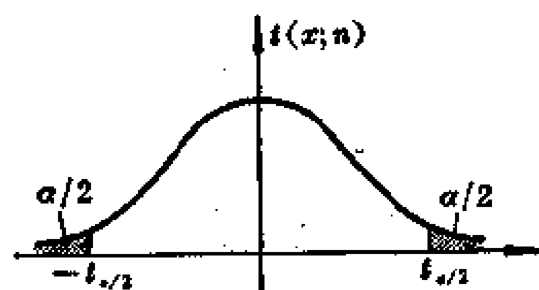


图 1

[例 3] 为了调查应用克矽平治疗矽肺的效果，今抽查应用克矽平治疗矽肺患者 10 名，记录下治疗前后血红蛋白的含量，如下表．试问，该药是否会引起血红蛋白含量的变化($\alpha=0.05$)？

克矽平治疗前后矽肺患者血红蛋白含量

病人号	血红蛋白(克%)		治疗前后差 x (克%)	x^2
	治疗前	治疗后		
1	11.3	14.0	2.7	7.29
2	15.0	13.8	−1.2	1.44
3	15.0	14.0	−1.0	1.00
4	13.5	13.5	0	0
5	12.8	13.5	0.7	0.49
6	10.0	12.0	2.0	4.00
7	11.0	14.7	3.7	13.69
8	12.0	11.4	−0.6	0.36
9	13.0	13.8	0.8	0.64
10	12.3	12.0	−0.3	0.09
Σ			6.8	29.00

我们假设治疗前后血红蛋白的差 X 服从正态分布 $N(\mu, \sigma^2)$，则问题归结为检验假设 $H_0: \mu=0$。

现在，由数据表计算得到

$$n=10$$

$$\bar{x}=0.68$$

$$s_n^2=\frac{1}{n}\sum_{i=1}^{10}(x_i-\bar{x})^2=\frac{1}{10}\left(29.00-\frac{(6.8)^2}{10}\right)=2.4376$$

$$t=\frac{\bar{x}-\mu_0}{s_n}\sqrt{n-1}=\frac{0.68}{\sqrt{2.4376}}\times 3\approx 1.31$$

当假设 H_0 为真时，统计量 T 服从自由度为 9 的 t-分布，对 $\alpha=0.05$，查 t-分布表(见附表 3)得到 $t_{\alpha/2}=2.262$. 今因 $|t|=1.31<2.262=t_{\alpha/2}$，所以不能拒绝假设 $H_0(\mu=0)$. 即我们不能否认"血红蛋白无变化"这一假定。

t-检验法还可以应用于比较两个带有未知方差(但方差相等)的正态母体的均值是否相等的问题.

设正态母体 X 和 Y，$X\sim N(\mu_1, \sigma^2)$，$Y\sim N(\mu_2,\sigma^2)$. 其中 σ^2，μ_1，μ_2 是未知参数，要求检验假设 $H_0: \mu_1=\mu_2$.

分别从 X，Y 两母体抽得容量为 m 和 n 的两个独立子样，其子样均值和方差分别记为

$$\bar{X}=\frac{1}{m}\sum_{i=1}^{m}X_i, \quad \bar{Y}=\frac{1}{n}\sum_{i=1}^{n}Y_i$$

$$S_{1m}^2=\frac{1}{m}\sum_{i=1}^{m}(X_i-\bar{X})^2$$

$$S_{2n}^2=\frac{1}{n}\sum_{i=1}^{n}(Y_i-\bar{Y})^2$$

与前面类似的理由，可构造统计量

$$T=\frac{\bar{X}-\bar{Y}}{\sqrt{mS_{1m}^2+nS_{2n}^2}}\sqrt{\frac{mn(m+n-2)}{m+n}} \tag{7}$$

按系 2.4.3，在 $H_0(\mu_1=\mu_2)$ 成立的条件下由(7)式定义的统

计量 T 服从 $t(m+n-2)$ 分布. 否则服从非中心 t–分布, 与前同样, 对给定的水平 α, 由 t–分布表查得 $t_{\alpha/2}$, 然后, 从观察数据计算出 T 的观察值 t, 比较 $|t|$ 和 $t_{\alpha/2}$, 若 $|t|>t_{\alpha/2}$, 则拒绝假设 H_0.

这里特别要注意的是要求母体 X 和 Y 的方差必需相等, 否则即使假设 $H_0(\mu_1=\mu_2)$ 成立, 也不能推得由(7)式决定的统计量 T 是服从 t–分布的.

[例 4] 为研究正常成年男、女血液红细胞的平均数之差别, 检查某地正常成年男子 156 名, 正常成年女子 74 名, 计算得男性红细胞平均数为 465.13 万/mm³, 子样标准差为 54.80 万/mm³; 女性红细胞平均数为 422.16 万/mm³, 子样标准差为 49.20 万/mm³. 试检验该地正常成年人的红细胞平均数是否与性别有关.

这个问题可以处理如下:

已知两个母体 X 表示正常成年男性的红细胞数, Y 表示正常成年女性的红细胞数. 由经验知道 X, Y 均服从正态分布, 且方差相同. 要求检验

$$H_0: \mu_1=\mu_2.$$

现在

$$m=156, \bar{x}=465.13 \text{ 万/mm}^3$$
$$s_{1m}=54.80 \text{ 万/mm}^3$$
$$n=74, \bar{y}=422.16 \text{ 万/mm}^3$$
$$s_{2n}=49.20 \text{ 万/mm}^3$$

代入公式(7)得

$$t=\frac{\bar{x}-\bar{y}}{\sqrt{ms_{1m}^2+ns_{2n}^2}}\sqrt{\frac{mn(m+n-2)}{m+n}}$$
$$=\frac{465.13-422.16}{\sqrt{156\times(54.80)^2+74\times(49.20)^2}}\sqrt{\frac{156\times74\times(156+74-2)}{156+74}}$$
$$\approx\frac{42.97}{7.49}\approx5.73$$

自由度 $n=m+n-2=156+74-2=228$

若取 $\alpha=0.01$ 得 $t_{\alpha/2}=2.601$ 于是 $|t|=5.73>2.601=t_{\alpha/2}$，所以拒绝假设 H_0，即认为正常成年男、女性红细胞数有显著差别.

二、方差的检验

对于正态母体，大家知道它是被两个参数——均值 μ 和方差 σ^2 所完全决定的，上面我们介绍了检验关于均值的假设 $H_0(\mu=\mu_0)$ 的 u–检验和 t–检验. 在许多实际问题中，常常要求检验关于方差的假设，在运用 t–检验比较两个正态母体的均值时，曾明确指出要求两母体有相等的方差，即 $D(\mathbb{X})=D(\mathbb{Y})=\sigma^2$（$\sigma^2$ 未知）. 由于 $D(\mathbb{X})$ 和 $D(\mathbb{Y})$ 是未知的，因此它们的相等除通过一般的直观分析之外，是否能构造一个检验方法检验它们是相等的假设呢？下面将介绍处理这类问题的检验方法.

设 $X_1, \cdots, X_n$ 是从正态母体 $\mathbb{X}$ 中抽得的一个子样，欲检验假设 H_0: $\sigma^2=\sigma_0^2$.

在§1.3 中已经看到子样方差 S_n^2 是 σ^2 的最大似然估计，并由定理 1.2.2 得到，$ES_n^2=\frac{n-1}{n}\sigma^4$，$D(S_n^2)=\frac{2(n-1)}{n^2}\sigma^4$，它们都与均值 μ 无关. 由此可见当 $H_0(\sigma^2=\sigma_0^2)$ 成立时，S_n^2 较集中在 $\frac{n-1}{n}\sigma_0^2$ 的周围波动，否则将偏离 $\frac{n-1}{n}\sigma_0^2$. 因此子样方差是构造检验假设 $H_0(\sigma^2=\sigma_0^2)$ 的合适的统计量. 为了查表便利，将它标准化得到

$$\chi^2=\frac{nS_n^2}{\sigma_0^2}=\frac{\sum_{i=1}^{n}(X_i-\overline{X})^2}{\sigma_0^2} \tag{8}$$

由定理 2.4.1 知道，由(8)式定义的统计量 χ^2 在假设 $H_0(\sigma^2=\sigma_0^2)$ 成立时，服从自由度为 $n-1$ 的 χ^2–分布.

对给定显著水平 α，我们选取 $c_{1\alpha}$ 和 $c_{2\alpha}$ 使得下面两式成立.

$$\int_{c_{1\alpha}}^{c_{2\alpha}}\chi^2(x;n-1)dx=1-\alpha \tag{9}$$

$$\int_{c_{1\alpha}}^{c_{2\alpha}} \chi^2(x; n-1)dx = \frac{1}{n-1}\int_{c_{1\alpha}}^{c_{2\alpha}} x\chi^2(x; n-1)dx \qquad (10)$$

将子样观察数据代入(8)式，计算出 χ^2 的观察值 $\hat{\chi}^2$. 如果 $\hat{\chi}^2 \leqslant c_{1\alpha}$ 或 $\hat{\chi}^2 \geqslant c_{2\alpha}$，则拒绝假设 $H_0(\sigma=\sigma_0)$，否则认为试验结果与假设无显著差异.

由于

$$\begin{aligned}
&\int_{c_{1\alpha}}^{c_{2\alpha}} [x-(n-1)]\chi^2(x; n-1)dx \\
&= \frac{1}{\Gamma\left(\frac{n-1}{2}\right)2^{\frac{n-1}{2}}}\int_{c_{1\alpha}}^{c_{2\alpha}} [x-(n-1)]x^{\frac{n-1}{2}-1}e^{-\frac{x}{2}}dx \\
&= \frac{1}{\Gamma\left(\frac{n-1}{2}\right)2^{\frac{n-1}{2}}}\left[-2x^{\frac{n-1}{2}}e^{-\frac{x}{2}}\Big|_{c_{1\alpha}}^{c_{2\alpha}}\right] \\
&= \frac{2}{\Gamma\left(\frac{n-1}{2}\right)2^{\frac{n-1}{2}}}\left[c_{1\alpha}^{\frac{n-1}{2}}e^{-c_{1\alpha}/2} - c_{2\alpha}^{\frac{n-1}{2}}e^{-c_{2\alpha}/2}\right]
\end{aligned}$$

所以(10)式又可等价地写为

$$c_{1\alpha}^{\frac{n-1}{2}}e^{-c_{1\alpha}/2} = c_{2\alpha}^{\frac{n-1}{2}}e^{-c_{2\alpha}/2} \qquad (11)$$

假设检验的理论指出，上述构造的检验法在某种意义下是最优的. 但是如此选取 $c_{i\alpha}$, $i=1, 2$ 很麻烦，不便于实际应用. 所以通常用下式选取 $c_{i\alpha}$, $i=1, 2$.

$$\int_0^{c_{1\alpha}} \chi^2(x; n-1)dx = \int_{c_{2\alpha}}^{\infty} \chi^2(x; n-1)dx = \alpha/2 \qquad (12)$$

(见图 2). 这儿的 $c_{i\alpha}, i=1, 2$ 可从附表 2 查得.

类似地，我们有时也需要比较两个正态母体的方差是否相等。

设 $X_1, X_2, \cdots, X_m$ 和 $Y_1, Y_2, \cdots, Y_n$ 是分别从服从 $N(\mu_1, \sigma_1^2)$ 和 $N(\mu_2, \sigma_2^2)$ 分布的两个母体中抽得的独立子样，现欲检验假设 $H_0: \sigma_1^2=\sigma_2^2$.

对此检验问题，我们采用 F-统计量

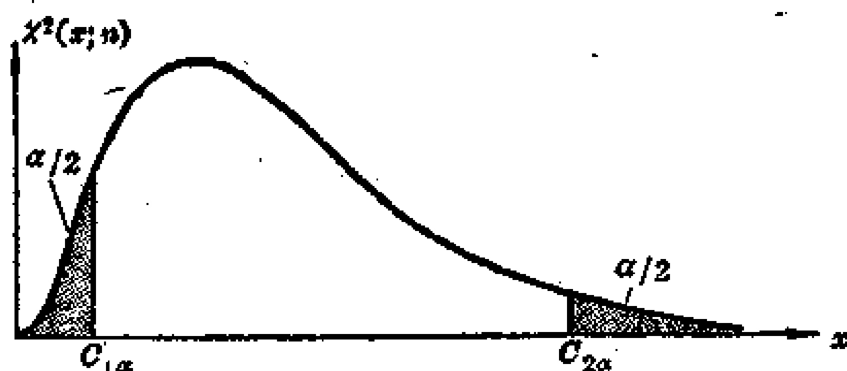


图 2

$$F=\frac{mS_{1m}^2}{nS_{2n}^2}\cdot\frac{n-1}{m-1} \tag{13}$$

其中 $S_{1m}^2=\frac{1}{m}\sum_{i=1}^{m}(X_i-\overline{X})^2$, $S_{2n}^2=\frac{1}{n}\sum_{i=1}^{n}(Y_i-\overline{Y})^2$.

由系 2.4.2 知道，(13) 式所决定的统计量 F 在假设 $H_0(\sigma_1^2=\sigma_2^2)$成立时,服从 $F(m-1, n-1)$分布. 又因为当假设 $H_0(\sigma_1^2=\sigma_2^2)$成立时,

$$E\left(\frac{m}{m-1}S_{1m}^2\right)=E\left(\frac{n}{n-1}S_{2n}^2\right)$$

而当 $\sigma_1^2>\sigma_2^2$ 时, 由于

$$E\left(\frac{m}{m-1}S_{1m}^2\right)=\sigma_1^2>\sigma_2^2=E\left(\frac{n}{n-1}S_{2n}^2\right)$$

而使得统计量 F 有偏大的趋势; 当 $\sigma_1^2<\sigma_2^2$ 时, 由于

$$E\left(\frac{m}{m-1}S_{1m}^2\right)<E\left(\frac{n}{n-1}S_{2n}^2\right)$$

而使得统计量 F 有偏小的趋势. 所以对于给定的显著水平 α, 通常是选取 $f_{1\alpha}$ 和 $f_{2\alpha}$, 使得

$$\int_0^{f_{1\alpha}} f(x; m-1, n-1)\,dx=\int_{f_{2\alpha}}^{\infty} f(x; m-1, n-1)dx$$
$$=\alpha/2 \tag{14}$$

(见图 3). 这里 $f_{1\alpha}$, $f_{2\alpha}$ 可从附表 4 查得. 然后运用子样观察值代入(13)式, 算得统计量 F 的观察值 f, 如果 $f\leqslant f_{1\alpha}$ 或 $f\geqslant f_{2\alpha}$ 则

拒绝 $H_0(\sigma_1^2=\sigma_2^2)$. 否则认为实验结果与假设 $H_0(\sigma_1^2=\sigma_2^2)$ 无显著差异. 由于这个检验法运用的统计量服从 F-分布, 所以常称为 F-检验法.

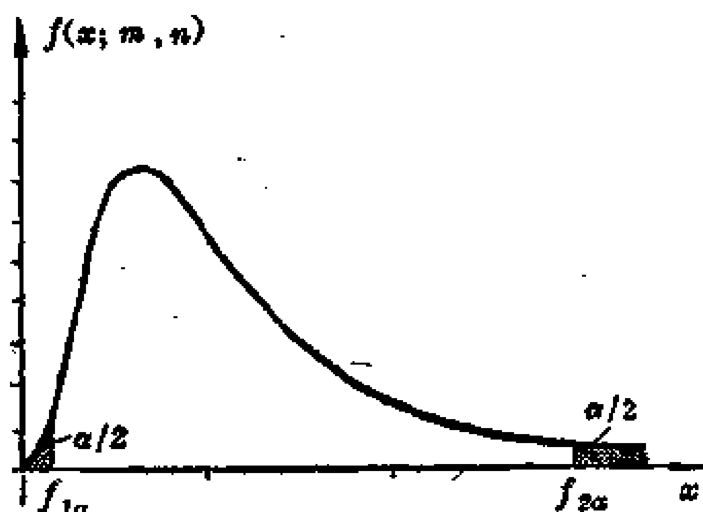


图 3

若记 $f_{m,n,\alpha}$ 是满足下面(15)式的 α-分位点,

$$\int_0^{f_{m,n,\alpha}} f(x;\ m,\ n)\,dx=\alpha \tag{15}$$

则由 F-分布的性质有

$$\int_{1/f_{m,n,\alpha}}^{\infty} f(x;\ n,\ m)\,dx=\alpha \tag{16}$$

即对于自由度 $(m-1,\ n-1)$ 的 $f_{2\alpha}$ 可以从自由度为 $(n-1,\ m-1)$ 的 $f_{1\alpha}$ 求得, 反之亦然.

[例 5] (续例 4) 在例 4 中, 我们假设男、女性红细胞数的分布的方差相等, 现在我们就来检验这一假设 H_0: $\sigma_1^2=\sigma_2^2$.

这里 $m=156$, $s_{1m}=54.80$, $n=74$, $s_{2n}=49.20$

代入(13)式得

$$f=\frac{156\times(54.80)^2}{74\times(49.20)^2}\times\frac{73}{155}=\frac{468474.24}{179127.36}\times\frac{73}{155}\approx1.25$$

自由度是(155, 73). 若给定 $\alpha=0.10$, 查表得

$$f_{1\alpha}\approx0.719,\quad f_{2\alpha}\approx1.43$$

现在 $$f_{1\alpha}=0.719<f=1.25<1.43=f_{2\alpha}$$

故我们不能拒绝假设 $H_0(\sigma_1^2=\sigma_2^2)$，即认为试验结果与假设无显著差异.

一般地，对于两个正态母体 X 和 Y，如果它们的方差是未知的而需比较它们的均值时，首先用 F-检验法检验它们的方差是否一致，如果检验结果是接受方差相等这一假设．然后再运用 t-检验方法，比较它们的均值.

现将上述所描写的检验方法列表如下.

表 1　正态母体参数的显著性检验

检验参数	假　设 H_0	统　计　量	分　布
μ	$\mu=\mu_0(\sigma=\sigma_0)$	$U=\frac{\bar{X}-\mu_0}{\sigma_0}\sqrt{n}$	$N(0,1)$
	$\mu_1=\mu_2(\sigma_1,\sigma_2$已知)	$U=(\bar{X}-\bar{Y})\Big/\sqrt{\frac{\sigma_1^2}{m}+\frac{\sigma_2^2}{n}}$	
	$\mu=\mu_0,\ \sigma^2>0$	$T=\frac{\bar{X}-\mu_0}{S_n}\sqrt{n-1}$	$t(n-1)$
	$\mu_1=\mu_2(\sigma_1=\sigma_2)$	$T=\frac{(\bar{X}-\bar{Y})}{\sqrt{mS_{1m}^2+nS_{2n}^2}}\cdot\sqrt{\frac{mn(m+n-2)}{m+n}}$	$t(m+n-2)$
σ^2	$\sigma=\sigma_0$	$\chi^2=\frac{nS_n^2}{\sigma_0^2}$	$\chi^2(n-1)$
	$\sigma_1^2=\sigma_2^2$	$F=\frac{mS_{1m}^2}{nS_{2n}^2}\cdot\frac{n-1}{m-1}$	$F(m-1,n-1)$

下面作为正态母体参数的显著性检验的应用，介绍 $(\bar{x}, R)$ 质量控制图.

三、$(\bar{x}, R)$质量控制图

在自动化生产的条件下，如何应用科学的方法控制每道（主要）工序的产品质量是个重要问题．大家知道，自动机床生产出的零件的尺寸，其均匀性是很好的，但也有差异，这在统计学上统称为误差．造成这种误差的因素很多，而按其性质可以分成两大类：

系统误差与随机误差．系统误差大多是由于某种具体的个别确定性的原因所造成．例如生产设备的缺陷；工艺规程不完善；没有遵守工艺规程；检查用具使用不当；原料、工具、夹具的缺陷等等，这类原因对产品的质量影响较大，能造成大量产品的不合格．除此之外，还有随机误差，它是由许多随机性因素（有的没有控制，有的无法控制）造成的，例如温湿度的变化；电网电压的波动；金属部件的不时伸缩；原料的不均匀；机械振动等等，它们使产品的尺寸在规定的标准尺寸的上下波动．由中心极限定理知道，这种波动近似服从正态分布．由于这种随机波动的存在，在产品的设计图纸上，不仅规定出标准的中心尺寸，而且也给定所允许的误差范围，即所谓“公差”．质量控制的目的就是要及时发现系统误差，分析造成系统误差的原因并及时调整．

由于产品被控制的尺寸（记为 X）的波动是由各种微小的随机因素的影响造成的，根据中心极限定理，可认为它服从正态分布 $N(\mu,\sigma^2)$，其中参数 μ 反映产品的平均尺寸，σ^2 反映尺寸 X 的波动的大小．因此生产的正常与否完全被这两个参数决定．所以生产过程中的质量控制问题就是要控制这两个参数始终能保持在正常值范围里，用数理统计的方法就可以定时地抽样检验 μ 和 σ^2 是否正常，如果有一个被检验为不正常就及时进行调整．所以这里可将质量控制问题归结为正态分布的 μ 和 σ^2 的检验问题．下面介绍的平均值控制图和极差控制图就是分别用来检验 μ 和 σ^2 的一种具体的实施方法．

1．平均值控制图

设 $x_1,\cdots,x_n$ 是在生产过程中抽取的一组容量为 n 的子样，计算出它的子样均值 $\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i$，由于大多数情况下生产是正常的，所以一般采用 3σ 检验准则．即若检验假设 $H_0\ (\mu=\mu_0)$，则当

$|\bar{x}-\mu_0|>3\sigma_{\bar{x}}(=3\sigma/\sqrt{n})$时拒绝 $H_0(\mu=\mu_0)$，否则认为试验结果与假设无显著差异，即接受假设 $H_0(\mu=\mu_0)$.

由正态分布的性质知道，如果 $\bar{X}$ 服从 $N(\mu_0, \sigma^2/n)$分布，则

$$P\left\{|\bar{X}-\mu_0|\geqslant\frac{3}{\sqrt{n}}\sigma\right\}=0.0027$$

这是一个小概率事件，约一千次中出现3次．所以这是一个相当于取水平 $\alpha=0.0027$ 的 u-检验法．

平均值控制图的作法如下(见图4)：

作控制上限和下限关于中心尺寸对称，其距离 $h=\frac{3}{\sqrt{n}}\sigma$，这相当于一个关于假设 H_0(μ=中心尺寸)的 $\alpha=0.0027$ 的 u-检验法，如果公差$\gg\frac{3}{\sqrt{n}}\sigma$，则不取 h，改取 $\delta=\left(3+\frac{3}{\sqrt{n}}\right)\sigma$，取(公差上限$-\delta$，公差下限$+\delta$)为控制限，这相当于检验假设 H_0(μ 落在区间(公差上限-3σ，公差下限$+3\sigma$)之内)的 $\alpha=0.0027$ 的 u-检验法．由正态分布性质知道，当 μ 落在这个区间(公差上限-3σ，公差下限$+3\sigma$)之中时，可以保证产品的尺寸有99.73%落在公差范围内，一般这是满意了．有时也取 $\delta=\left(1.2+\frac{3}{\sqrt{n}}\right)\sigma$．其中 σ 是未知的，可用 $\hat{\sigma}$ 代替，$\hat{\sigma}$ 的具体计算方法在后面讲．

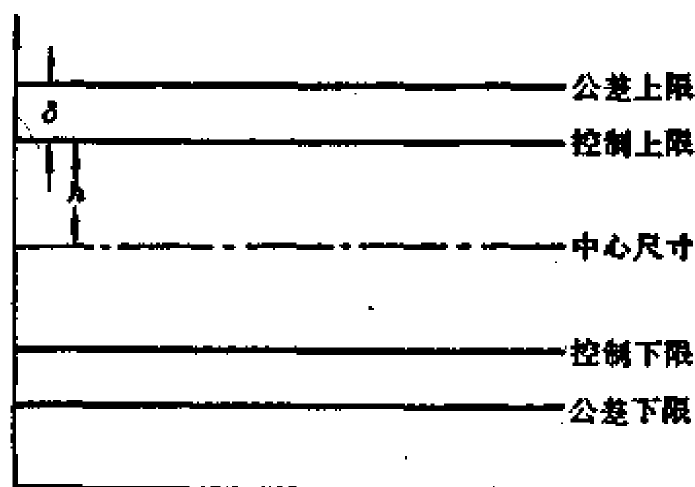


图4 平均值控制图

2. 极差控制图

由于计算子样方差比较繁琐，日常运用，不受欢迎，故一般用极差 $R_n = x_n^{(n)} - x_n^{(1)} = \max(x_1, \cdots, x_n) - \min(x_1, \cdots, x_n)$ 来代替（这里应取 $n \leqslant 10$）. 直观上讲方差 σ^2 大，离散程度大，则 R 也较大；反之亦然. 容易证明，对于正态分布的母体，$E(R)$ 仅依赖于 σ. 它们之间的具体关系是

$$E(R_n) = c_n \sigma \tag{17}$$

c_n 的具体数值由下表给出

n	2	3	4	5	6	7	8	9	10
c_n	1.128	1.693	2.059	2.326	2.534	2.704	2.847	2.970	3.078

R_n 的根方差 σ_R 与 σ 之间也有类似关系

$$\sigma_R = \sqrt{D(R_n)} = d_n \cdot \sigma \tag{18}$$

d_n 的具体数值是：

n	2	3	4	5	6	7	8	9	10
d_n	0.853	0.888	0.880	0.864	0.848	0.833	0.820	0.808	0.797

由(17)、(18)式得到

$$\sigma_R = d_n \cdot \sigma = \frac{d_n}{c_n} E(R_n) \tag{19}$$

极差控制图的作法(见图 5).

同样是采用 3σ 准则，其中

$$\text{控制上限} = \bar{R} + 3\hat{\sigma}_R = \bar{R} + 3\frac{d_n}{c_n}\bar{R} = \left(1 + 3\frac{d_n}{c_n}\right)\bar{R}$$

$$\text{控制下限} = \bar{R} - 3\hat{\sigma}_R = \left(1 - 3\frac{d_n}{c_n}\right)\bar{R}$$

如果上式右端的值小于零，则取为零，其中 $\bar{R}$ 是 $E(R_n)$ 的估计值。

如果 $1-3\frac{d_n}{c_n}<0$，那么控制下限就可以取消，也许有人要问：离散程度当然是愈小愈好，何必要下限呢？回答是：工厂希望 R 愈小愈好，但 R 太小可能是由于使用的原料质量过高所致，这对工厂未必合算，另外也可能是由于测量仪器不灵敏所致，这就得检查仪器。所以控制下限是有用的。

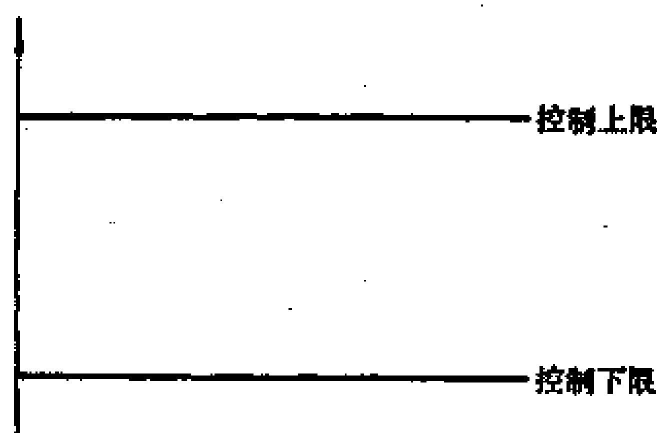


图 5　极差控制图

使用时，平均值控制图与极差控制图通常联合运用，每隔一定时间（如 1 小时）抽取几个产品观察，记录下数据，算出 $\bar{x}_n$ 和 R_n，如两者均各自落在控制限之内，则认为正常，否则需进行检查。

控制图的使用不仅能及时发现问题，及时调整，保证生产正常。而且通过长期的资料、经验积累，还可以摸索出一套分析造成越出控制限的原因的方法，对生产设备的维修节约大量的人力、物力。

最后，简单地介绍一下此处获得 $\bar{R}$ 和 $\hat{\sigma}$ 的常用方法。

在生产正常的情况下，每隔一定时间（如一小时）抽取 n 个（一般取 $n=5$）产品进行观察，得一组（n 个）数据，总共抽取 l 组（$l=20\sim32$）数据，计算出每组的极差 R_i，然后计算 R_i 的平均值

$$\bar{R}=\frac{1}{l}\sum_{i=1}^{l}R_i \tag{20}$$

再运用(17)式算得

$$\hat{\sigma}=\frac{1}{c_n}\bar{R} \tag{21}$$

§3. 正态母体参数的置信区间

在例 2.4 中我们通过 t-检验发现男、女性的红细胞数有显著差异，并从子样均值看到，男性的红细胞平均数比女性的要高，但到底高多少，在许多实际问题中，我们往往希望通过子样的观察给出一个范围，使得这个范围能按足够大的概率(给定的)包含我们所感兴趣的参数. 在统计学上，我们称这个范围叫**置信区间**(或**置信域**)，这类问题称为**区间估计问题**.

参数的置信区间与参数的假设检验之间有着密切的联系，下面我们通过例 2.4 来介绍构造置信区间的方法，并说明这种联系.

若设 $EX=\mu_1$(男性的红细胞平均数)

$EY=\mu_2$(女性的红细胞平均数)

作

$$T=\frac{[(\bar{X}-\mu_1)-(\bar{Y}-\mu_2)]}{\sqrt{mS_{1m}^2+nS_{2n}^2}}\sqrt{\frac{mn(m+n-2)}{m+n}} \tag{1}$$

显然，$T\sim t(m+n-2)$. 与§2 提到的 t-检验法一样，对给定的水平 α，由 t-分布表可查得 $t_{\alpha/2}$ 使

$$P\{|T|<t_{\alpha/2}\}=1-\alpha \tag{2}$$

解不等式 $|T|<t_{\alpha/2}$ 得

$$(\bar{X}-\bar{Y})-\frac{t_{\alpha/2}\sqrt{mS_{1m}^2+nS_{2n}^2}\cdot\sqrt{m+n}}{\sqrt{(m+n-2)mn}}<(\mu_1-\mu_2)$$

$$<(\bar{X}-\bar{Y})+\frac{t_{\alpha/2}\sqrt{mS_{1m}^2+nS_{2n}^2}\cdot\sqrt{m+n}}{\sqrt{(m+n-2)mn}} \tag{3}$$

记

$$\hat{\theta}_1\triangleq(\bar{X}-\bar{Y})-\frac{t_{\alpha/2}\sqrt{mS_{1m}^2+nS_{2n}^2}\cdot\sqrt{m+n}}{\sqrt{(m+n-2)mn}} \tag{4}$$

$$\hat{\theta}_2 \triangleq (\overline{X}-\overline{Y})+\frac{t_{\alpha/2}\sqrt{mS_{1m}^2+nS_{2n}^2}\cdot\sqrt{m+n}}{\sqrt{(m+n-2)mn}} \tag{5}$$

则由(2)式～(5)式，得到

$$P\{\hat{\theta}_1<\mu_1-\mu_2<\hat{\theta}_2\}=1-\alpha \tag{6}$$

(6)式说明，随机区间$(\hat{\theta}_1, \hat{\theta}_2)$以概率$1-\alpha$包含参数$(\mu_1-\mu_2)$，我们称$(\hat{\theta}_1, \hat{\theta}_2)$为$(\mu_1-\mu_2)$的置信区间，概率$(1-\alpha)$称为置信度。

把上面的结果应用于例2.4. 在此例中，

$$m=156, \quad \bar{x}=465.13, \quad s_{1m}=54.80$$
$$n=74, \quad \bar{y}=422.16, \quad s_{2n}=49.20$$

对$\alpha=0.01$，查表得$t_{\alpha/2}=2.601$，将这些数据代入(4)式和(5)式，可算得对置信度为0.99的$\mu_1-\mu_2$的置信区间是(23.49, 62.45)，即男性的红细胞平均数约比女性的高23.49万/mm³到62.45万/mm³.

类似地，可以从§2的正态母体参数的各种检验法构造正态母体参数的各种置信区间(见下表1)。

下面我们讨论正态分布均值和方差的联合置信域。

(μ, σ^2)的联合置信域可以运用$\overline{X}$和S_n^2的联合分布来构造。由定理2.4.1知道，$\overline{X}$和S_n^2是独立的，因此，如果我们希望寻找置信水平为0.95的置信域，我们可以找到数a，和c_1, c_2，使得

$$P\left\{-a<\frac{\overline{X}-\mu_0}{\sigma_0/\sqrt{n}}<a\right\}=\sqrt{0.95}\approx 0.975 \tag{7}$$

和

$$P\left\{c_1<\frac{nS_n^2}{\sigma_0^2}<c_2\right\}=\sqrt{0.95}\approx 0.975 \tag{8}$$

联合概率是

$$P\left\{-a<\frac{\overline{X}-\mu_0}{\sigma_0/\sqrt{n}}<a,\ c_1<\frac{nS_n^2}{\sigma_0^2}<c_2\right\}=0.95 \tag{9}$$

解得：

表 1　正态母体参数的置信区间

待估参数	条件		置信区间下限 θ_1	置信区间上限 θ_2	对应的检验统计量
μ	单子样	$\sigma=\sigma_0$	$\overline{X}-u_{\alpha/2}\cdot\sigma_0/\sqrt{n}$	$\overline{X}+u_{\alpha/2}\cdot\sigma_0/\sqrt{n}$	$U=\frac{\overline{X}-\mu}{\sigma_0}\sqrt{n}$
		σ 未知	$\overline{X}-t_{\alpha/2}\cdot S_n/\sqrt{n-1}$	$\overline{X}+t_{\alpha/2}\cdot\frac{S_n}{\sqrt{n-1}}$	$T=\frac{\overline{X}-\mu}{S_n}\sqrt{n-1}$
$\mu_1-\mu_2$	双子样	已知 $\sigma_1=\sigma_2$ 但数值未知	$(\overline{X}-\overline{Y})-t_{\alpha/2}\frac{\sqrt{mS_{1m}^2+nS_{2n}^2}\cdot\sqrt{m+n}}{\sqrt{(m+n-2)mn}}$	$(\overline{X}-\overline{Y})+t_{\alpha/2}\frac{\sqrt{mS_{1m}^2+nS_{2n}^2}\cdot\sqrt{m+n}}{\sqrt{(m+n-2)mn}}$	$T=\frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{\sqrt{mS_{1m}^2+nS_{2n}^2}}\times\sqrt{\frac{(m+n-2)m\cdot n}{m+n}}$
σ^2	单子样		$\frac{1}{C_{2\alpha}}\cdot nS_n^2$	$\frac{1}{C_{1\alpha}}\cdot nS_n^2$	$\chi^2=\frac{nS_n^2}{\sigma_0^2}$
$\frac{\sigma_1^2}{\sigma_2^2}$	双子样		$\frac{1}{f_{2\alpha}}\cdot\frac{mS_{1m}^2}{nS_{2n}^2}\cdot\frac{n-1}{m-1}$	$\frac{1}{f_{1\alpha}}\cdot\frac{mS_{1m}^2}{nS_{2n}^2}\cdot\frac{n-1}{m-1}$	$F=\frac{mS_{1m}^2/\sigma_1^2}{nS_{2n}^2/\sigma_2^2}\cdot\frac{n-1}{m-1}$

$$P\left\{(\mu_0-\bar{X})^2<a^2\sigma_0^2/n,\ \frac{nS_n^2}{c_2}<\sigma_0^2<\frac{nS_n^2}{c_1}\right\}=0.95 \qquad (10)$$

由此可见，$(\mu,\ \sigma^2)$的置信度为 0.95 的联合置信域是 (10) 式大括号内不等式对 μ_0, σ_0^2 所给出的范围，它由图 1 中的阴影部分所示。

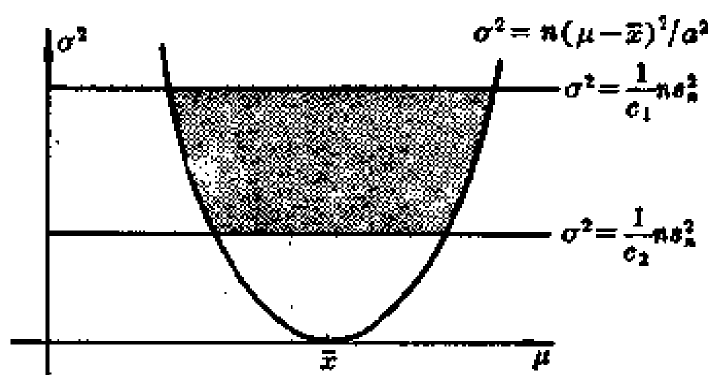


图 1

§4. 多项分布的 χ^2–检验

前面介绍了正态母体参数的显著性检验问题，在这些检验问题中，先假定母体的分布是已知的，具有正态分布. 然后在这前提下，根据从该母体中抽得的子样对分布的未知均值或方差的假设进行检验. 然而在许多场合，事先不知道母体分布的类型. 此时首先需要根据子样对母体分布的种种假设进行检验. 本节所介绍的多项分布的 χ^2–检验就是其中的一种方法.

一、分布的 χ^2–检验法

设母体 X 是仅取 r 个可能值的离散型随机变量，不失一般性，设 X 的可能取值是 1, 2, …, r，记它取值 i 的概率为 p_i，即

$$P(X=i)=p_i,\ i=1,\ \cdots,\ r$$

和

$$\sum_{i=1}^{r}p_i=1$$

设 $X=(X_1, \cdots, X_n)$ 是从该母体中所抽得的简单随机子样，$(x_1, \cdots, x_n)$ 是子样观察值. 记 n_i 表示 $(x_1, \cdots, x_n)$ 中取值为 i 的个数，即子样中出现事件 $(X=i)$ 的频数，则得到 $n_1, \cdots, n_r$. 显然，每个 n_i 都是子样的函数，所以它们也是随机变量，且有

$$\sum_{i=1}^{r} n_i = n$$

和 $(n_1, \cdots, n_r)$ 服从多项分布，它的概率分布是

$$\frac{n!}{n_1!n_2!\cdots n_r!} p_1^{n_1} \cdot p_2^{n_2} \cdots p_r^{n_r} \tag{1}$$

我们知道，频率是概率的反映，如果母体的概率分布确实是 $(p_1, \cdots, p_r)$ 那末当观察个数 n 愈来愈大时，频率 n_i/n 与 p_i 之间的差异将愈来愈小. 因此频率 n_i/n 和 p_i 之间的差异程度可以反映出 $(p_1, \cdots, p_r)$ 是否为母体的真分布. 皮尔逊(K. Pearson) 首先提出运用下面的统计量来衡量它们的差异程度，

$$\chi^2 = \sum_{i=1}^{r} \frac{(n_i - np_i)^2}{np_i} \tag{2}$$

这个统计量又称为皮尔逊统计量.

直观上比较清楚，如果 $(p_1, \cdots, p_r)$ 是母体服从的真实概率分布，统计量 χ^2 要偏小些，否则它就有偏大的趋势. 因此它可以用来作为多项分布的检验统计量. 但是我们还需要知道它的分布. 下面的皮尔逊定理给出了它的渐近分布.

定理 1 （皮尔逊定理）

当 $(p_1, \cdots, p_r)$ 是母体的真实概率分布时，由(2)式定义的统计量 χ^2 的渐近分布是自由度为 $r-1$ 的 χ^2-分布. 即其渐近分布具有下式给出的分布密度函数

$$\chi^2(x; r-1) = \begin{cases} \dfrac{1}{2^{\frac{r-1}{2}} \Gamma\left(\dfrac{r-1}{2}\right)} x^{\frac{r-3}{2}} e^{-\frac{x}{2}}, & \text{当 } x>0 \\ 0, & x<0 \end{cases} \tag{3}$$

[证明] 因$(n_1, \cdots, n_r)$服从多项分布，它的联合概率分布由(1)式给出，所以它的特征函数是

$$\varphi_{1n}(t_1, \cdots, t_r) = (p_1 e^{it_1} + \cdots + p_r e^{it_r})^n \tag{4}$$

记

$$Y_i \triangleq \frac{n_i - np_i}{\sqrt{np_i}}, \quad i = 1, \cdots, r \tag{5}$$

则

$$\sum_{i=1}^{r} Y_i \sqrt{p_i} = 0$$

$$\chi^2 = \sum_{i=1}^{r} \frac{(n_i - np_i)^2}{np_i} = \sum_{i=1}^{r} Y_i^2$$

而$(Y_1, \cdots, Y_r)$的联合分布的特征函数是

$$\varphi_n(t_1, \cdots, t_r) = e^{-i\sqrt{n}\sum_{j=1}^{r} t_j\sqrt{p_j}} \left(p_1 e^{\frac{it_1}{\sqrt{np_1}}} + \cdots + p_r e^{\frac{it_r}{\sqrt{np_r}}}\right)^n$$

$$\log \varphi_n(t_1, \cdots, t_r) = -i\sqrt{n} \sum_{j=1}^{r} t_j \sqrt{p_j} + n \log\left(\sum_{j=1}^{r} p_j e^{\frac{it_j}{\sqrt{np_j}}}\right) \tag{6}$$

利用泰勒展开

$$e^{\frac{it_j}{\sqrt{np_j}}} - 1 = \frac{it_j}{\sqrt{np_j}} - \frac{t_j^2}{2np_j} + o\left(\frac{1}{n}\right)$$

和

$$\log(1+x) = x - \frac{x^2}{2} + o(x^3)$$

有

$$\begin{aligned}\log \varphi_n(t_1, \cdots, t_r) = & -i\sqrt{n} \sum_{j=1}^{r} t_j \sqrt{p_j} + n\left\{ i \sum_{j=1}^{r} \frac{t_j\sqrt{p_j}}{\sqrt{n}} \right.\\ & \left. - \frac{1}{2n}\sum_{j=1}^{r} t_j^2 - \frac{1}{2}\left(i \sum_{j=1}^{r} \frac{t_j\sqrt{p_j}}{\sqrt{n}}\right)^2 \right\} + o(1) \\ & \xrightarrow[n\to\infty]{} -\frac{1}{2}\left\{\sum_{j=1}^{r} t_j^2 - \left(\sum_{j=1}^{r} t_j \sqrt{p_j}\right)^2\right\}\end{aligned}$$

即

$$\begin{aligned}\lim_{n\to\infty} \varphi_n(t_1, \cdots, t_r) &= \varphi(t_1, \cdots, t_r) \\ &= \exp\left\{-\frac{1}{2}\left[\sum_{j=1}^{r} t_j^2 - \left(\sum_{j=1}^{r} t_j \sqrt{p_j}\right)^2\right]\right\}\end{aligned}$$

作正交变换：

$$\begin{cases} Z_k=\sum_{j=1}^{r} a_{kj}Y_j,\ k=1,\ \cdots,\ r-1 \\ Z_r=\sum_{j=1}^{r}\sqrt{p_j}\cdot Y_j \end{cases}$$

和
$$\begin{cases} u_k=\sum_{j=1}^{r} a_{kj}t_j,\ k=1,\ \cdots,\ r-1 \\ u_r=\sum_{j=1}^{r}\sqrt{p_j}\, t_j \end{cases}$$

则
$$\sum_{j=1}^{r} t_j^2-\left(\sum_{j=1}^{r} t_j\sqrt{p_j}\right)^2=\sum_{j=1}^{r-1} u_j^2$$

当 $n\to\infty$ 时 $(Z_1,\ \cdots,\ Z_r)$ 的特征函数 $\varphi_n^{(Z)}(u_1,\ \cdots,\ u_r)$ 趋近于

$$\varphi^{(Z)}(u_1,\ \cdots,\ u_r)=\exp\left\{-\frac{1}{2}\sum_{j=1}^{r-1} u_j^2\right\}$$

这表明 $Z_1,\ \cdots,\ Z_{r-1}$ 的分布收敛于相互独立的正态 $N(0,\ 1)$ 分布，Z_r 依概率收敛于 0，所以

$$\chi^2=\sum_{j=1}^{r} Y_j^2=\sum_{j=1}^{r} Z_j^2$$

渐近于自由度为 $r-1$ 的 χ^2–分布.

由定理 1，当子样容量 n 足够大时，可以近似地认为由(2)式定义的 χ^2 统计量服从于自由度为 $r-1$ 的 χ^2–分布. 因此当我们要检验某个假设 H_0: $p_i=p_{i0}$, $i=1,\ \cdots,\ r$ 时，只要将 p_{i0} 代入(2)式，根据子样观察计算出 χ^2 统计量的观察值

$$\hat{\chi}^2=\sum_{i=1}^{r}\frac{(n_i-np_{i0})^2}{np_{i0}} \tag{7}$$

并对于给定的水平 α，查 χ^2–分布表(附表 2)求出临界值 χ_α^2，比较 $\hat{\chi}^2$ 和 χ_α^2，如果 $\hat{\chi}^2\geqslant\chi_\alpha^2$ 则作出拒绝假设 H_0 的决定．认为试验结果与假设有显著差异.

皮尔逊统计量也可用来检验母体是否服从某个给定的分布函数 $F_0(x)$．问题的叙述和作法如下．

设子样 $X=(X_1,\cdots,X_n)$是取自分布为 $F(x)$的一个简单随机子样. 欲检验假设 H_0: $F(x)=F_0(x)$. ($F_0(x)$是某个已知的分布)

为此我们选取 $r-1$ 个实数 $-\infty<y_1<y_2<\cdots<y_{r-1}<+\infty$ 它们将实数轴区分为 r 个区间 $(-\infty, y_1)$, $[y_1, y_2)$, $[y_2, y_3)$, $\cdots$, $[y_{r-1}, +\infty)$, 记

$$
\begin{aligned}
&p_1 \triangleq F(y_1) \\
&p_i \triangleq F(y_i)-F(y_{i-1}), \qquad i=2, \cdots, r-1 \qquad (8)\\
&p_r \triangleq 1-F(y_{r-1})
\end{aligned}
$$

并用 n_1 表示子样观察值落在$(-\infty, y_1)$中的个数, n_i 表示子样观察值落在区间$[y_{i-1}, y_i)$中的个数, n_r 表示子样观察值落在$[y_{r-1}, +\infty)$中的个数. 则 $n_1, \cdots, n_r$ 服从多项分布. 当假设 H_0: $F(x)=F_0(x)$成立时, 由定理 1 得到, 统计量

$$\chi^2=\sum_{i=1}^{r}\frac{(n_i-np_{i0})^2}{np_{i0}} \qquad (9)$$

的分布渐近于自由度为 $r-1$ 的 χ^2-分布. 因此可以按上述方法对假设 $H_0(F(x)=F_0(x))$进行检验, (9)式中的 p_{i0} 是将 F_0 代替(8)式中的 F 算得的 p_i 值.

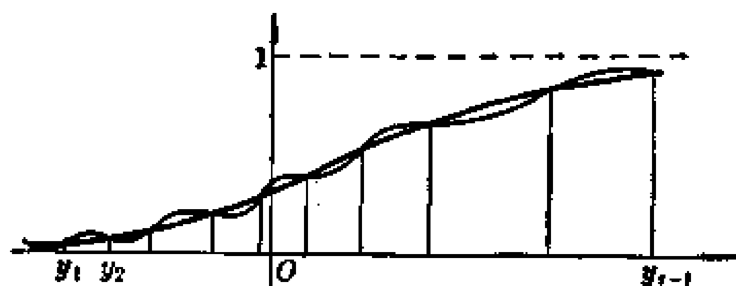


图 1

在图 1 中我们可以看到有时虽 $F(x)\neq F_0(x)$, 但是在某一种分划之下, 可有

$$F(y_i)-F(y_{i-1})=F_0(y_i)-F_0(y_{i-1})=p_{i0}, \qquad i=2, \cdots, r-1$$

和
$$F(y_1)=F_0(y_1)=p_{10}$$

此时虽然 $F(x) \neq F_0(x)$，但在这种分划之下，它并不影响由(9)式给出的 χ^2 统计量的数值，因而很容易把不真的假设也接受过来。不过因为我们的分组是随机的，而且一般分布函数较为光滑，即使相差一些也不会太大，所以也就很少考虑到这一问题。这里也可看到采用这一检验法实际上只是检验了假设

$$H_0: F(y_i) = F_0(y_i), \quad i=1, \cdots, r-1$$

而并未真正去检验母体的分布是否为 $F_0(x)$。关于分布函数的其它一些检验法在第七章中再介绍。

[例 1] 在某黑盒中存放有白球和黑球。现作下面这样的试验：用返回抽取方式从此黑盒中摸球，直到摸取的是白球为止，记录下抽取的次数。重复执行如此的试验 100 次，其结果如下：

抽取次数	1	2	3	4	≥5
频　　数	43	31	15	6	5

试问该黑盒中的白球与黑球的个数是否相等($\alpha=0.05$)。

[解] 记随机变量 X 表示首次出现白球所需的摸取次数，则 X 服从几何分布

$$P\{X=k\} = (1-p)^{k-1}p, \qquad k=1, 2, \cdots$$

其中 p 表示从此黑盒中任意摸取一球，出现是白球的概率。

如果黑盒中白球与黑球的个数相等，此时 $p=\frac{1}{2}$，代入上式得到

$$P\{X=1\} = \frac{1}{2}, \qquad P\{X=2\} = \frac{1}{4}$$

$$P\{X=3\} = \frac{1}{8}, \qquad P\{X=4\} = \frac{1}{16}$$

$$P\{X \geq 5\} = \sum_{k=5}^{\infty} 2^{-k} = \frac{1}{16}$$

现作假设

$$H_0:\ p_1=\frac{1}{2},\ \ p_2=\frac{1}{4},\ \ p_3=\frac{1}{8},\ \ p_4=\frac{1}{16},\ \ p_5=\frac{1}{16}$$

将此及试验结果(频数)代入(7)式，得到 χ^2 统计量的观察值是

$$\hat{\chi}^2=\sum_{k=1}^{5}\frac{(n_k-np_k)^2}{np_k}=\frac{(43-50)^2}{50}+\frac{(31-25)^2}{25}+\frac{(15-12.5)^2}{12.5}+\frac{(6-6.25)^2}{6.25}+\frac{(5-6.25)^2}{6.25}=3.2$$

对 $\alpha=0.05$，自由度 $=5-1=4$，由 χ^2-分布表查得 $\chi^2_\alpha=9.488$，现在 $3.2<9.488$，因此认为试验结果与假设无显著差异，即认为黑盒中白球与黑球个数相等.

在上面的讨论中，我们考虑了假设 $H_0(F(x)=F_0(x))$ 完全指定了分布 $F_0(x)$ 的情形. 但在许多实际场合，假设 H_0 只给出分布的具体函数形式，而其中包含有未知参数. 例如最常遇见的是要检验“母体是否服从正态分布”这一假设，这里假设 H_0 给出一个分布类 $N(\mu,\sigma^2)$，其中包含有两个未知参数 μ 和 σ^2. 现在由于不能从(8)式计算出 p_{i0}，因此(9)式不能计算，上述检验方法就不能直接运用，须进行修改. 怎样解决这一问题呢？一个自然的做法是在(8)式的计算中用估计量代替未知参数.

令 $\hat{\theta}_1,\cdots,\hat{\theta}_s$ 是未知参数 $\theta_1,\cdots,\theta_s$ 的最大似然估计，将它代入(8)式得到

$$\hat{p}_i=F_0(y_i;\hat{\theta}_1,\cdots,\hat{\theta}_s)-F_0(y_{i-1};\hat{\theta}_1,\cdots,\hat{\theta}_s),\quad i=2,\cdots,r-1 \tag{10}$$

$$\hat{p}_1=F_0(y_1;\hat{\theta}_1,\cdots,\hat{\theta}_s)$$

$$\hat{p}_r=1-F_0(y_{r-1};\hat{\theta}_1,\cdots,\hat{\theta}_s)$$

将此代入(9)式得

$$\chi^2=\sum_{i=1}^{r}\frac{(n_i-n\hat{p}_i)^2}{n\hat{p}_i} \tag{11}$$

则由(11)式给定的统计量渐近于自由度为 $r-s-1$ 的 χ^2-分布①. 它可以用来检验包含有未知参数的分布假设.

[例 2] 下表给出了某地 120 名 12 岁男孩身高的资料.

表 1 某地 120 名 12 岁男孩身高资料(单位:厘米)

128.1	144.4	150.3	146.2	140.6	139.7
134.1	124.3	147.9	143.0	143.1	142.7
126.0	125.6	127.7	154.4	142.7	141.2
133.4	131.0	125.4	130.3	146.3	146.8
142.7	137.6	136.9	122.7	131.8	147.7
135.8	134.8	139.1	139.0	132.3	134.7
138.4	136.6	136.2	141.6	141.0	138.4
145.1	141.4	139.9	140.6	140.2	131:0
150.4	142.7	144.3	136.4	134.5	132.3
152.7	148.1	139.6	138.9	136.1	135.9
140.3	137.3	134.6	145.2	128.2	135.9
140.2	136.6	139.5	135.7	139.8	129.1
141.4	139.7	136.2	138.4	138.1	132.9
142.9	144.7	318.8	138.3	135.3	140.6
142.2	152.1	142.4	142.7	136.2	135.0
154.3	147.9	141.3	143.8	138.1	139.7
127.4	146.0	155.8	141.2	146.4	139.4
140.8	127.7	150.7	160.3	148.5	147.5
138.9	123.1	126.0	150.0	143.7	156.9
133.1	142.8	136.8	133.1	144.5	142.4

用 X 表示 12 岁男孩的身高, 要求利用 χ^2-检验法来检验假设 H_0: X 服从正态分布.

这里由于假设没有给出 X 的均值和方差, 而仅说它服从正态分布. 因此需先求出它们的最大似然估计. 它们的最大似然估计分别是(见例 1.3.7)

① 这一结论的证明和 $F(x;\theta_1,\cdots,\theta_s)$ 需满足的正则性条件可参阅克拉美,《统计学数学方法》,上海科学技术出版社,1966,第一版.

$$\hat{\mu}=\frac{1}{n}\sum_{i=1}^{n}X_i$$

$$\hat{\sigma}^2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2$$

根据表 1 中的数据算得

$$\hat{\mu}=\bar{x}\approx 139.5$$

$$\hat{\sigma}^2=s_n^2\approx 55$$

$$\hat{\sigma}\approx 7.42$$

在 χ^2 检验中，一般要求分组时每组中的观察个数不少于 5 个. 在表 1 中看到，这批数据的最小值是 122.7，最大值是 160.3. 现选用 8 个实数 126, 130, 134, 138, 142, 146, 150, 154 将实数轴分为 9 个区间，得到各组的频数如下：

组　号	1	2	3	4	5
区间界限	(−∞, 126)	[126, 130)	[130, 134)	[134, 138)	[138, 142)
频　数	5	8	10	22	33

组　号	6	7	8	9
区间界限	[142, 146)	[146, 150)	[150, 154)	[154, ∞)
频　数	20	11	6	5

下面计算相应的 $\hat{p}_i$, $i=1, 2, \cdots, 9$.

$$\hat{p}_1=P(X<126|\hat{\mu},\hat{\sigma})=P\left(\frac{X-\hat{\mu}}{\hat{\sigma}}<\frac{126-\hat{\mu}}{\hat{\sigma}}\right)$$

$$=\Phi\left(\frac{126-\hat{\mu}}{\hat{\sigma}}\right)=\Phi\left(\frac{126-139.5}{7.42}\right)$$

$$=\Phi(-1.82)=0.0344$$

$$\hat{p}_2 = P(126 < X < 130 \mid \hat{\mu}, \hat{\sigma})$$

$$= \Phi\left(\frac{130-139.5}{7.42}\right) - \Phi\left(\frac{126-139.5}{7.42}\right)$$

$$= \Phi(-1.28) - \Phi(-1.82)$$

$$= 0.1003 - 0.0344 = 0.0669$$

$$\hat{p}_3 = \Phi\left(\frac{134-139.5}{7.42}\right) - \Phi\left(\frac{130-139.5}{7.42}\right)$$

$$= \Phi(-0.74) - \Phi(-1.28) = 0.2297 - 0.1003$$

$$= 0.1294$$

$$\hat{p}_4 = \Phi\left(\frac{138-139.5}{7.42}\right) - \Phi\left(\frac{134-139.5}{7.42}\right)$$

$$= \Phi(-0.20) - \Phi(-0.74)$$

$$= 0.4207 - 0.2297 = 0.1910$$

$$\hat{p}_5 = \Phi\left(\frac{142-139.5}{7.42}\right) - \Phi\left(\frac{138-139.5}{7.42}\right)$$

$$= \Phi(0.34) - \Phi(-0.20) = 0.6331 - 0.4207$$

$$= 0.2124$$

$$\hat{p}_6 = \Phi\left(\frac{146-139.5}{7.42}\right) - \Phi\left(\frac{142-139.5}{7.42}\right)$$

$$= \Phi(0.88) - \Phi(0.34) = 0.8106 - 0.6331$$

$$= 0.1775$$

$$\hat{p}_7 = \Phi\left(\frac{150-139.5}{7.42}\right) - \Phi\left(\frac{146-139.5}{7.42}\right)$$

$$= \Phi(1.42) - \Phi(0.88) = 0.9222 - 0.8106$$

$$= 0.1116$$

$$\hat{p}_8 = \Phi\left(\frac{154-139.5}{7.42}\right) - \Phi\left(\frac{150-139.5}{7.42}\right)$$

$$= \Phi(1.95) - \Phi(1.42) = 0.9744 - 0.9222$$

$$= 0.0522$$

$$\hat{p}_9=1-\Phi\left(\frac{154-139.5}{7.42}\right)=1-\Phi(1.95)$$

$$=1-0.9744=0.0256$$

下面计算 $\hat{\chi}^2$ 的值

组　　号	1	2	3	4	5
n_i	5	8	10	22	33
$n\hat{p}_i$	4.128	8.028	15.53	22.92	25.49
$n_i-n\hat{p}_i$	0.872	-0.028	-5.53	-0.92	7.51
$(n_i-n\hat{p}_i)^2$	0.7604	0.0008	30.5809	0.8464	56.4001
$(n_i-n\hat{p}_i)^2/n\hat{p}_i$	0.1817	0.0001	1.9659	0.0369	2.2126

组　　号	6	7	8	9
n_i	20	11	6	5
$n\hat{p}_i$	21.30	13.39	6.26	3.07
$n_i-n\hat{p}_i$	-1.30	-2.39	-0.26	1.93
$(n_i-n\hat{p}_i)^2$	1.69	5.7121	0.0676	3.7249
$(n_i-n\hat{p}_i)^2/n\hat{p}_i$	0.0793	0.4265	0.0108	1.2133

所以
$$\hat{\chi}^2=\sum_{i=1}^{9}\frac{(n_i-n\hat{p}_i)^2}{n\hat{p}_i}=6.1271$$

自由度是 $9-2-1=6$，对 $\alpha=0.05$，得 $\chi_\alpha^2=12.952$，显然 $\hat{\chi}^2=6.1271<12.952=\chi_\alpha^2$. 因此我们认为试验结果与正态分布这一假设无显著差异.

二、联立表的独立性检验

假定考察一个二元母体，或者说考察母体中诸元素的两个指标 (X, Y). 将这两个指标的取值范围分别分成 r 和 q 个互不相交的区间 $A_1, \cdots, A_r$ 和 $B_1, \cdots, B_q$. 设从该母体中抽取一个容量为 n 的子样 $(x_1, y_1), \cdots, (x_n, y_n)$. 用 n_{ij} 表示子样观察中其 x 指标

落于 A_i 而其 y 指标落于 B_j 中的个数($i=1, \cdots, r; j=1, \cdots, q$).
又记

$$n_{i\cdot} \triangleq \sum_{j=1}^{q} n_{ij}, \quad n_{\cdot j} \triangleq \sum_{i=1}^{r} n_{ij} \tag{12}$$

显然

$$n = \sum_{i=1}^{r} \sum_{j=1}^{q} n_{ij} \tag{13}$$

我们用下表来表示子样元素的这种分类(这张表称为联立表).

表 2　二元联立表

		B_j				$\sum (n_{i\cdot})$
		1	2	$\cdots$	q	
A_i	1	n_{11}	n_{12}	$\cdots$	n_{1q}	$n_{1\cdot}$
	2	n_{21}	n_{22}	$\cdots$	n_{2q}	$n_{2\cdot}$
	$\vdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\vdots$
	r	n_{r1}	n_{r2}	$\cdots$	n_{rq}	$n_{r\cdot}$
$\sum (n_{\cdot j})$		$n_{\cdot 1}$	$n_{\cdot 2}$	$\cdots$	$n_{\cdot q}$	n

我们提出假设 H_0: 母体元素的两个指标 $\mathbb{X}$ 和 $\mathbb{Y}$ 是相互独立的. 如果记

$$p_{ij} \triangleq P(\mathbb{X} \in A_i, \mathbb{Y} \in B_j), \ i=1, \cdots, r; \ j=1, \cdots, q$$

和

$$p_{i\cdot} \triangleq P(\mathbb{X} \in A_i), \ i=1, \cdots, r$$
$$p_{\cdot j} \triangleq P(\mathbb{Y} \in B_j), \ j=1, \cdots, q \tag{14}$$

显然有

$$p_{i\cdot} = \sum_{j=1}^{q} p_{ij}$$
$$p_{\cdot j} = \sum_{i=1}^{r} p_{ij} \tag{15}$$

$$\sum_{i=1}^{r} p_{i\cdot} = \sum_{j=1}^{q} p_{\cdot j} = 1 \tag{16}$$

而在假设 H_0 成立的条件下, 有 $p_{ij} = p_{i\cdot} p_{\cdot j}$, 所以联立表中的独立性

检验就是检验假设 H_0: $p_{ij}=p_{i\cdot}p_{\cdot j}$, $i=1, \cdots, r$; $j=1, \cdots, q$.

在这个假设中没有明确指出 $r+q$ 个未知参数 $p_{i\cdot}$ 和 $p_{\cdot j}$ 的值. 由(16)式, 其中仅有 $r+q-2$ 个独立的未知参数. 要想用 χ^2 检验(11)式来检验假设 H_0, 需先求出这些未知参数的最大似然估计. 这里最大似然函数是

$$L=\prod_{i=1}^{r}\prod_{j=1}^{q}p_{ij}^{n_{ij}}=\prod_{i=1}^{r}\prod_{j=1}^{q}(p_{i\cdot}p_{\cdot j})^{n_{ij}}=\prod_{i=1}^{r}p_{i\cdot}^{n_{i\cdot}}\prod_{j=1}^{q}p_{\cdot j}^{n_{\cdot j}}$$

$$=\left(1-\sum_{i=1}^{r-1}p_{i\cdot}\right)^{n_{r\cdot}}\left(1-\sum_{j=1}^{q-1}p_{\cdot j}\right)^{n_{\cdot q}}\prod_{i=1}^{r-1}p_{i\cdot}^{n_{i\cdot}}\prod_{j=1}^{q-1}p_{\cdot j}^{n_{\cdot j}}$$

$$\log L=n_{r\cdot}\log\left(1-\sum_{i=1}^{r-1}p_{i\cdot}\right)+n_{\cdot q}\log\left(1-\sum_{j=1}^{q-1}p_{\cdot j}\right)$$

$$+\sum_{i=1}^{r-1}n_{i\cdot}\log p_{i\cdot}+\sum_{j=1}^{q-1}n_{\cdot j}\log p_{\cdot j}$$

解似然方程组

$$\frac{\partial\log L}{\partial p_{i\cdot}}=\frac{-n_{r\cdot}}{1-\sum_{i=1}^{r-1}p_{i\cdot}}+\frac{n_{i\cdot}}{p_{i\cdot}}=\frac{n_{i\cdot}}{p_{i\cdot}}-\frac{n_{r\cdot}}{p_{r\cdot}}=0$$

$$i=1, \cdots, r-1$$

$$\frac{\partial\log L}{\partial p_{\cdot j}}=\frac{-n_{\cdot q}}{1-\sum_{j=1}^{q-1}p_{\cdot j}}+\frac{n_{\cdot j}}{p_{\cdot j}}=\frac{n_{\cdot j}}{p_{\cdot j}}-\frac{n_{\cdot q}}{p_{\cdot q}}=0$$

$$j=1, \cdots, q-1$$

得到

$$\hat{p}_{i\cdot}=\frac{n_{i\cdot}}{n},\quad i=1, \cdots, r;\ \hat{p}_{\cdot j}=\frac{n_{\cdot j}}{n},\quad j=1, \cdots, q \tag{17}$$

代入公式(11)得到统计量

$$\chi^2=n\sum_{i=1}^{r}\sum_{j=1}^{q}\frac{\left(n_{ij}-\frac{n_{i\cdot}n_{\cdot j}}{n}\right)^2}{n_{i\cdot}n_{\cdot j}} \tag{18}$$

在这里通过似然方程解出 $r+q-2$ 参数, 所以自由度是

$$rq-(r+q-2)-1=(r-1)(q-1)$$

因此当 $H_0(p_{ij}=p_{i\cdot}p_{\cdot j})$ 成立时, 由(18)式决定的 χ^2 统计量的分布

渐近于自由度为$(r-1)(q-1)$的χ^2-分布．它是我们判断H_0所需要的检验统计量．

当$r=q=2$时，联立表2称为是四分的（又称为四格表）．

［例3］ 调查339名50岁以上吸烟习惯与患慢性气管炎病的关系，得下表．试问吸烟者与不吸烟者慢性气管炎患病率是否有所不同．

表 3

	患慢性气管炎者	未患慢性气管炎者	合　计	患病率
吸　烟	43	162	205	21.0
不吸烟	13	121	134	9.7
$\sum$	56	283	339	16.5

在这里对每个对象考察两个指标，X——是否吸烟；Y——是否患慢性气管炎．它们均各取两种值，A_1——吸烟，A_2——不吸烟；B_1——患慢性气管炎，B_2——未患慢性气管炎．所以$r=q=2$．

由(17)式得

$$\hat{p}_{1\cdot}=\frac{205}{339},\qquad \hat{p}_{2\cdot}=\frac{134}{339}$$

$$\hat{p}_{\cdot 1}=\frac{56}{339},\qquad \hat{p}_{\cdot 2}=\frac{283}{339}$$

代入(18)式得

$$\begin{aligned}\hat{\chi}^2&=\frac{(43-205\times 56/339)^2}{205\times 56/339}+\frac{(162-205\times 283/339)^2}{205\times 283/339}\\&\quad+\frac{(13-134\times 56/339)^2}{134\times 56/339}+\frac{(121-134\times 283/339)^2}{134\times 283/339}\\&=\frac{(43-33.86)^2}{33.86}+\frac{(162-171.14)^2}{171.14}+\frac{(13-22.14)^2}{22.14}\\&\quad+\frac{(121-111.86)^2}{111.86}=7.48\end{aligned}$$

自由度$=(r-1)(q-1)=1$，对 $\alpha=0.01$ 查表(附表2)得 $\chi^2_\alpha=6.635$，显然 $7.48>6.635$，所以我们拒绝假设 H_0，即认为慢性气管炎的患病率与吸烟有关.

§5. 广义似然比检验

在§2和§4的讨论中，我们从分析具体问题的直观背景出发，针对不同的问题给出了不同的检验方法. 这一节是在前面讨论的基础上提出一个构造检验方法的较为一般的法则.

在§2和§4中看到，对于一个参数的显著性检验问题，先假设母体 X 服从某一分布 $F(x;\theta)$，其中 θ 是未知参数. 然后对 θ 提出某种假设 $H_0: \theta=\theta_0$ 要求检验. 这实际上意味着 X 的真分布可能服从由异于 θ_0 的值 θ' 对应的分布函数 $F(x;\theta')$. 类似于参数估计问题，我们把母体 X 可能取的 θ 值的全体记为 Θ，它在假设检验里称为容许假设. 因此 $\{F(x;\theta),\ \theta\in\Theta\}$ 实际上是母体 X 的可能分布族. 当我们拒绝假设 H_0 时，实际上就是认为 X 的真分布属于分布族 $\{F(x;\theta),\ \theta\in\Theta-\theta_0\}$. 现将参数的显著性检验问题表述如下.

设 $X=(X_1,\ \cdots,\ X_n)$ 是从母体 X 中抽取的子样，X 的可能分布族是 $\{f(x;\theta),\ \theta\in\Theta\}$，其中 θ(可以是向量)是未知参数(当 X 是连续型变量时 f 表示分布密度，当 X 是离散型变量时 f 表示概率分布.)要求检验假设 $H_0: \theta=\theta_0$. 这里应指出，θ_0 有时是表示一个集合，如在运用 t-检验法检验假设 $H_0: \mu=\mu_0$ 时，那里

$$\theta \triangleq (\mu,\ \sigma^2)$$

$$\Theta=\{(\mu,\ \sigma^2):\ -\infty<\mu<\infty,\ \sigma^2>0\}$$

$$\theta_0 \triangleq \{(\mu_0,\ \sigma^2):\ \sigma^2>0\}$$

它是一个未知参数的集合而不是一个单点.

现在我们引进一个统计量：

$$\lambda(x) \triangleq \frac{\sup\limits_{\theta\in\theta_0} \prod\limits_{i=1}^{n} f(x_i;\theta)}{\sup\limits_{\theta\in\Theta} \prod\limits_{i=1}^{n} f(x_i;\theta)} \tag{1}$$

习惯上称 $\lambda(x)$ 为广义似然比．显然它是子样的函数，不依赖于未知参数 θ．由于 $\theta_0 \subset \Theta$，所以

$$0 \leqslant \lambda(x) \leqslant 1 \tag{2}$$

类似于最大似然原理，如果 $\lambda(x)$ 取值较小，这说明当 H_0 为真时观察到样点 x 的概率比 H_0 不真时观察到样点 x 的概率要小得多，此时我们有理由怀疑假设 H_0 不真．所以从广义似然比出发，该检验问题是当下式成立时拒绝 H_0，

$$\lambda(x) \leqslant \lambda_0 \tag{3}$$

其中 λ_0 的选取是使得下式成立，

$$P_\theta\{\lambda(X) \leqslant \lambda_0\} \leqslant \alpha, \quad 对一切 \theta \in \theta_0 \tag{4}$$

由(3)、(4)式给出的检验法称为水平为 α 的**广义似然比检验**．当 θ_0 是一个单点时(4)式可写为

$$P_{\theta_0}\{\lambda(X) \leqslant \lambda_0\} \leqslant \alpha \tag{5}$$

［例 1］ 设母体 X 服从正态 $N(\mu, \sigma^2)$ 分布，其中 μ 和 σ^2 均是未知参数．欲检验假设 $H_0: \mu = \mu_0$．

设 $X=(X_1, \cdots, X_n)$ 是从该母体中抽得的一个简单随机子样，它的分布密度是

$$f(x; \mu, \sigma^2) = \left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^n \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right\}$$

在这里，参数空间是

$$\Theta = \{(\mu, \sigma^2): -\infty < \mu < \infty, \sigma^2 > 0\}$$

假设 H_0 所对应的子集是

$$\theta_0 = \{(\mu_0, \sigma^2): \sigma^2 > 0\}$$

运用微分法对 $f(x; \mu, \sigma^2)$ 求最大值得到

$$\sup_{(\mu,\sigma^2)\in\theta_0} f(x;\mu,\sigma^2)=\left[2\pi\cdot\frac{1}{n}\sum(X_i-\mu_0)^2\right]^{-n/2}e^{-n/2}$$

$$\sup_{(\mu,\sigma^2)\in\Theta} f(x;\mu,\sigma^2)=\left[2\pi\frac{1}{n}\sum(x_i-\bar{x})^2\right]^{-n/2}\ e^{-n/2}$$

所以广义似然比是

$$\lambda(x)=\frac{\sup\limits_{(\mu,\sigma^2)\in\theta_0} f(x;\mu,\sigma^2)}{\sup\limits_{(\mu,\sigma^2)\in\Theta} f(x;\mu,\sigma^2)}=\left\{\frac{\sum(x_i-\bar{x})^2}{\sum(x_i-\mu_0)^2}\right\}^{n/2}$$

$$=\left\{\frac{1}{1+n(\bar{x}-\mu_0)^2/\sum(x_i-\bar{x})^2}\right\}^{n/2}=\left\{\frac{1}{1+t^2/n-1}\right\}^{n/2} \qquad (6)$$

其中

$$t(x)=\frac{\sqrt{n(n-1)}\,(\bar{x}-\mu_0)}{\sqrt{\sum(x_i-\bar{x})^2}} \qquad (7)$$

所以广义似然比检验的是当下式成立时拒绝 H_0,

$$\lambda(x)<\lambda_0 \qquad (8)$$

由于 $\lambda(x)$ 是关于 $|t(x)|$ 的单调下降函数，所以(8)又可等价地表示为

$$|t(x)|>c_0 \qquad (9)$$

因为 $t(X)$ 当 H_0 成立时服从 $t(n-1)$ 分布，所以 c_0 由下式决定

$$\int_{c_0}^{\infty} t(x;n-1)dx=\alpha/2$$

它就是§2中所指出的 t-检验法.

[例 2] 设母体 X 是仅取 r 个值的离散型随机变量，它的概率分布是

$$P(X=i)=p_i,\quad i=1,\cdots,r$$

其中 $\theta=(p_1,\cdots,p_r)$ 是未知参数，满足关系式

$$\sum_{i=1}^{r}p_i=1$$

所以参数空间 Θ 是 $r-1$ 维.

欲检验假设 H_0: $p_i=p_{i0}$, $i=1,\cdots,r$. 现在 θ_0 是一个单点(p_{10},

$\cdots, p_{r0})$.

设 $X=(X_1, \cdots, X_n)$是从该母体中抽得的一个子样，分别以 $n_1, \cdots, n_r$ 表示子样观察 $x=(x_1, \cdots, x_n)$中取值为 1, …, r 的个数. 则

$$n_1+\cdots+n_r=n$$

$$P_\theta(X=x)=p_1^{n_1}\cdots p_r^{n_r}$$

因此

$$P_{\theta_0}(X=x)=p_{10}^{n_1}\cdots p_{r0}^{n_r}$$

而

$$\sup_{\theta\in\Theta} P_\theta(X=x)=\left(\frac{n_1}{n}\right)^{n_1}\cdots\left(\frac{n_r}{n}\right)^{n_r}$$

由此得到广义似然比是

$$\lambda(x)=\frac{P_{\theta_0}(X=x)}{\sup\limits_{\theta\in\Theta} P_\theta(X=x)}=\prod_{i=1}^{r}\left(\frac{np_{i0}}{n_i}\right)^{n_i}$$

由于

$$\begin{aligned}\log\lambda(x)&=\sum_{i=1}^{r} n_i\log\frac{np_{i0}}{n_i}\\&=-\sum_{i=1}^{r}[np_{i0}+(n_i-np_{i0})]\log\left(1+\frac{n_i-np_{i0}}{np_{i0}}\right)\\&=-\sum_{i=1}^{r}[np_{i0}+(n_i-np_{i0})]\left[\frac{n_i-np_{i0}}{np_{i0}}\right.\\&\qquad\left.-\frac{1}{2}\left(\frac{n_i-np_{i0}}{np_{i0}}\right)^2+\text{高阶项}\right]\\&=-\frac{1}{2}\sum_{i=1}^{r}\frac{(n_i-np_{i0})^2}{np_{i0}}+\text{高阶项}\end{aligned}$$

所以

$$-2\log\lambda(x)\approx\sum_{i=1}^{r}\frac{(n_i-np_{i0})^2}{np_{i0}}$$

又因为$-2\log\lambda(x)$关于$\lambda(x)$是单调下降的函数，所以广义似然比检验是当下式成立时拒绝 H_0,

$$\lambda(x)<\lambda_0$$

它可等价地写为

$$-2\log\lambda(x)>c$$

它又可近似地写为

$$\sum_{i=1}^{r}\frac{(n_i-np_{i0})^2}{np_{i0}}>c_0$$

这就是 § 3.4 中讨论过的 χ^2-检验.

习　　题

1. 某切割机在正常工作时，切割每段金属棒的平均长度为 10.5cm，标准差是 0.15cm. 今从一批产品中随机地抽取 15 段进行测量，其结果如下(单位：cm)：

10.4, 10.6, 10.1, 10.4, 10.5, 10.3, 10.3, 10.2
10.9, 10.6, 10.8, 10.5, 10.7, 10.2, 10.7

试问该机工作是否正常？($\alpha=0.05$)

2. 现有两箱灯泡，今从第一箱中抽取 9 只灯泡进行测试，得到它的平均寿命是 1532 小时，标准差是 423 小时；从第二箱中抽取 18只进行测试，得到它的平均寿命是 1412 小时，标准差是 380 小时. 作适当的假设，对水平 $\alpha=0.05$，检验是否可以认为这两箱灯泡是同一批生产的？

3. 用两种方法研究冰溶解时的潜热，对冷却到 −0.72℃ 的样品执行方法 I 和方法 II 两种处理. 下面给出了从 −0.72℃ 的冰溶解为 0℃ 的水的热量改变的数据(卡/克).

方法 I: 79.98, 80.04, 80.02, 80.04, 80.03, 80.03, 80.04, 79.97
80.05, 80.03, 80.02, 80.00, 80.02

方法 II: 80.02, 79.94, 79.98, 79.97, 80.03, 79.95, 79.97
79.97

试问对水平 $\alpha=0.05$，这两种处理方法的平均性能有否显著差异？

4. 为研究矽肺患者肺功能的变化情况，某医院对 I，II 期矽肺患者各 33 名测定其肺活量，得到 I 期患者的平均数为 2710 毫升，标准差为 147 毫升；II 期患者的平均数为 2830 毫升，标准差为 118 毫升. 对水平 $\alpha=0.05$，试问，第 I、II 期矽肺患者的肺活量有无显著差异？

5. 有一正四面体，将此四面体的四面分别涂为红，黄，蓝，白四种不同的颜色. 现作如下的抛掷试验：任意地抛掷该四面体，直到白色的一面与地面相接触为止，记录下抛掷的次数. 作如此的试验 200 次，其结果如下：

抛掷次数	1	2	3	4	≥5
频　　数	56	48	32	28	36

试问该四面体是否均匀($\alpha=0.05$)．

6．有一放射性物质，今在 2608 个相等长的时间间隔下进行观察(每个间隔为 7.5 秒)．记录下每个时间间隔内落于计数器中的质点个数．用 N_m 表示所计质点数为 m 的时间间隔总数，得下表(见第一册例 2.4.8)．

m	N_m	m	N_m
0	57	6	273
1	203	7	139
2	383	8	45
3	525	9	27
4	532	≥10	16
5	408		

试检验 m 是否服从普阿松分布($\alpha=0.05$)?

7．为研究慢性气管炎与吸烟量的关系，调查了 272 人，结果由下表给出．

	吸 烟 量 (支/日)			Σ
	0～	10～	20～	
患者人数	22	98	25	145
健康者人数	22	89	16	127
Σ	44	187	41	272

试问慢性气管炎与吸烟量是否有关($\alpha=0.05$)?

8．1972 年调查郊区某桑场采桑员和辅助工桑毛虫皮炎发病情况，结果如下表：

	采　桑	不采桑	合　计
患者人数	18	12	30
健康人数	4	78	82
合　　计	22	96	112

试问发生皮炎是否与工种有关？

9．设 $X \sim b(1, p)$，$0<p<1$，$(X_1, \cdots, X_n)$ 是一子样，试求检验问题 H_0: $p=p_0$ 的广义似然比检验．

10．设 $X \sim N(\mu, \sigma^2)$，$-\infty<\mu<\infty$，$\sigma^2>0$ 均是未知参数，$X_1, \cdots, X_n$ 是一子样，试求检验问题 H_0: $\sigma^2=\sigma_0^2$ 的广义似然比检验．

11．设 $X \sim N(\mu_1, \sigma_1^2)$，$Y \sim N(\mu_2, \sigma_2^2)$，$(X_1, \cdots, X_m)$ 和 $(Y_1, \cdots, Y_n)$ 是分别从中抽取的独立子样．试求检验问题 H_0: $\dfrac{\sigma_1^2}{\sigma_2^2}=\Delta_0$ 的广义似然比检验．

第四章　线性统计推断

§1. 最小二乘法

一、线性模型

在许多实际问题中，经常要求利用观察数据拟合系统的真实模型，其中最常见和最简单的系统模型是线性模型．如

[例 1]　一个作匀速直线运动的质点在时刻 t 的位置 S，它可用公式 $S=\alpha+\beta t$ 来描写，其中 α 是质点在 $t=0$ 时刻的初始位置，β 是平均速度．如果 α 和 β 是未知的，S 可以被观察，显然只要在两个不同的时刻观察质点的位置 S，就可以解出 α 和 β 的值．现假定由于某种原因，质点的位置不能被精确测量而是带有随机的测量误差．也就是说，被观察的不是 S 而是 $Y=S+e$，其中 e 是测量误差，它是一个均值为 0 的随机变量．代替 S 我们有公式

$$Y=S+e=\alpha+\beta t+e \tag{1}$$

其中 Y 是可观察的随机变量，t 是可观察的非随机量，e 是不可观察的随机变量，α 和 β 是未知参数．显然，现在通过在不同时刻对 Y 的观察，一般不能解出 α 和 β．而我们的目的是要求利用 t，Y 的各组观察值 $(t_i, Y_i)_{i=1,\cdots,n}$，用统计方法估计 α,β 和 S 在各个时刻的值．

由(1)得到观察值 (t_i, Y_i) 之间满足关系式

$$Y_i=\alpha+\beta t_i+e_i,\quad i=1,\ \cdots,n$$

用向量、矩阵形式表示，得

$$Y=X\beta+e \tag{2}$$

其中

$$Y \triangleq \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}, \quad X \triangleq \begin{pmatrix} 1 & t_1 \\ \vdots & \vdots \\ 1 & t_n \end{pmatrix}, \quad \beta \triangleq \begin{pmatrix} \alpha \\ \\ \beta \end{pmatrix}, \quad e \triangleq \begin{pmatrix} e_1 \\ \vdots \\ e_n \end{pmatrix}$$

[例 2]　在考察某化学反应时，发现反应的速度主要决定于所用催化剂的用量 x 及所加的温度 T，当 x 和 T 改变时反应速度也跟着改变. 但是由于观察条件及其它条件的影响，当 x 和 T 固定时，所观察到的反应速度也不完全相同. 即反应速度的观察值 Y 与 x, T 之间存在着关系式

$$Y = f(x, T) + e$$

其中 e 是不可观察的随机误差，我们的目的是通过试验——在不同的 x, T 值的条件下，观察反应速度 Y 的值，找出 $f(x, T)$ 的具体形式，从而通过控制 x 和 T 以达到控制反应速度. 由于 e 是不可观察的随机变量，所以这儿只能用统计方法.

为处理的方便起见，一般近似地取 f 是未知参数的线性函数，例如

$$f = \beta_1 x + \beta_2 T + \beta_0$$

(或更复杂的函数关系，如 $f = \beta_0 + \beta_1 x + \beta_2 T + \beta_3 x^2 + \beta_4 T^2 + \beta_5 xT$，等)

此时

$$Y = \beta_0 + \beta_1 x + \beta_2 T + e$$

设我们进行 n 次观察，得 $(Y_i, x_i, T_i)_{i=1,\cdots,n}$，则有

$$Y = X\beta + e \tag{3}$$

其中

$$Y \triangleq \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}, \quad X \triangleq \begin{pmatrix} 1 & x_1 & T_1 \\ \cdots & \cdots & \cdots \\ 1 & x_n & T_n \end{pmatrix}, \quad \beta \triangleq \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix}, \quad e \triangleq \begin{pmatrix} e_1 \\ \vdots \\ e_n \end{pmatrix}$$

例 1 和例 2 都把问题归结为要求通过 Y 估计未知参数 β. (2)式(或(3)式)就是通常所说的线性模型.

现把上述模型作一般的描述如下：

设因变量 Y 和自变量 $X_1, \cdots, X_m$ 服从线性关系

$$Y=\beta_1X_1+\beta_2X_2+\cdots+\beta_mX_m+e \tag{4}$$

$(Y_i, x_{i1}, x_{i2}, \cdots, x_{im})$, $i=1, 2, \cdots, n$ 是$(Y, X_1, \cdots, X_m)$的 n 个观察，它们满足关系

$$Y_i=\beta_1x_{i1}+\beta_2x_{i2}+\cdots+\beta_mx_{im}+e_i, \quad i=1,\cdots,n \tag{5}$$

记

$$Y\triangleq\begin{pmatrix}Y_1\\ \vdots\\ Y_n\end{pmatrix},\ X\triangleq\begin{pmatrix}x_{11} & \cdots & x_{1m}\\ & \cdots & \\ x_{n1} & \cdots & x_{nm}\end{pmatrix},\ \beta\triangleq\begin{pmatrix}\beta_1\\ \vdots\\ \beta_m\end{pmatrix},\ e\triangleq\begin{pmatrix}e_1\\ \vdots\\ e_n\end{pmatrix}$$

则(5)式可表示为

$$Y=X\beta+e \tag{6}$$

其中 X 是已知的 $n\times m$ 常数矩阵，β 是未知参数向量；e 是均值为零的随机向量，满足条件

$$Ee=0 \tag{7}$$

$$\operatorname{cov}(e, e)=\sigma^2I_n$$

I_n 是 $n\times n$ 单位阵，σ^2 是未知参数.

习惯上，我们称上述的 Y 服从线性模型，简记为 $(Y, X\beta, \sigma^2I_n)$.

由(6)式和(7)式，容易计算得

$$EY=X\beta, \quad \operatorname{cov}(Y, Y)=\sigma^2I_n \tag{8}$$

或写为

$$EY_i=\sum_{j=1}^{m}x_{ij}\beta_j$$

$$\operatorname{cov}(Y_i, Y_j)=\begin{cases}\sigma^2, & 当\ i=j,\\ 0, & 当\ i\neq j,\end{cases} \quad i, j=1, \cdots, n$$

对于由(6)，(7)式所确定的线性模型$(Y, X\beta, \sigma^2I_n)$，通常所考虑的统计推断问题是：对未知参数 β 和 σ^2 进行估计；对关于 β 的某种假设进行检验；对 Y 进行预报等等.

在下面的讨论中，如不作特殊说明，总假定 $n>m$ 和

$\mathrm{rank}(X)=m$①.

二、β 的估计

对于线性模型 $(Y, X\beta, \sigma^2 I_n)$，常采用最小二乘法寻找未知参数 β 的估计量 $\hat{\beta}$，它要求 β 的估计 $\hat{\beta}$ 满足下面的条件

$$\|Y-X\hat{\beta}\|^2=\min_{\beta}\|Y-X\beta\|^2 \tag{9}$$

其中 $\|Y-X\beta\|^2 \triangleq \sum_{i=1}^{n}\left(Y_i-\sum_{j=1}^{m}x_{ij}\beta_j\right)^2$，min 是对所有的 m 维向量 β 取极小值. 所以(9)式又可写为

$$\sum_{i=1}^{n}\left(Y_i-\sum_{j=1}^{m}x_{ij}\hat{\beta}_j\right)^2=\min_{\beta_1,\cdots,\beta_m}\sum_{i=1}^{n}\left(Y_i-\sum_{j=1}^{m}x_{ij}\beta_j\right)^2 \tag{10}$$

满足(10)式(或(9)式)的估计 $\hat{\beta}_L=\begin{pmatrix}\hat{\beta}_1\\ \vdots\\ \hat{\beta}_m\end{pmatrix}$ 称为 β 的最小二乘估计，这种求估计量的方法称为最小二乘法.

下面用微分法求满足(10)式的解 $\hat{\beta}_L$. 记

$$f(\beta_1, \cdots, \beta_m)\triangleq\sum_{i=1}^{n}\left(Y_i-\sum_{j=1}^{m}x_{ij}\beta_j\right)^2$$

令
$$\frac{\partial f}{\partial\beta_k}=0, \quad k=1, \cdots, m$$

得方程组
$$\sum_{i=1}^{n}\left(Y_i-\sum_{j=1}^{m}x_{ij}\beta_j\right)x_{ik}=0,\ k=1, \cdots, m$$

即
$$\sum_{i=1}^{n}Y_i x_{ik}=\sum_{i=1}^{n}\sum_{j=1}^{m}x_{ij}x_{ik}\beta_j=\sum_{j=1}^{m}\left(\sum_{i=1}^{n}x_{ij}x_{ik}\right)\beta_j,\ k=1, \cdots, m$$

用矩阵、向量形式表示，上述方程组又可写为

$$X^{\tau}Y=(X^{\tau}X)\beta \tag{11}$$

(11)式称为**正规方程**. 由于 X 的秩为 m，所以 $X^{\tau}X$ 是正定阵，存在逆阵 $(X^{\tau}X)^{-1}$，解方程(11)，得到 β 的最小二乘估计是

① 关于 X 是降秩时的详细讨论可参阅 (Rao, Linear Statistical Inference and Its Applications,(第二版)，第4章.

$$\hat{\beta}_L=(X^\tau X)^{-1}X^\tau Y \tag{12}$$

将 $\hat{\beta}_L$ 代入方程(4)，由于 e 一般是表示模型误差不可预测，略去 e 就得到因变量 Y 和自变量 $X_1, \cdots, X_m$ 之间的经验关系式

$$\hat{Y}=\hat{\beta}_1X_1+\hat{\beta}_2X_2+\cdots+\hat{\beta}_mX_m \tag{13}$$

其中

$$\hat{\beta}_L^\tau=(\hat{\beta}_1, \cdots, \hat{\beta}_m)$$

(13) 式常称为经验模型或经验回归方程． 当自变量取值为 $(X_1, \cdots, X_m)$ 时，它可以用来预测因变量 Y 的平均值 $\hat{Y}$.

下面讨论最小二乘估计 $\hat{\beta}_L$ 的一些基本性质：

性质 1 $\hat{\beta}_L$ 是 Y 的线性函数．

在统计里，如果估计量是子样的线性函数，则称它为**线性估计**．这里 $\hat{\beta}_L$ 是一线性估计．

性质 2 $\hat{\beta}_L$ 是 β 的无偏估计，因为

$$\begin{aligned} E\hat{\beta}_L&=E[(X^\tau X)^{-1}X^\tau Y] \\ &=(X^\tau X)^{-1}X^\tau\cdot EY \\ &=(X^\tau X)^{-1}X^\tau X\beta=\beta \end{aligned} \tag{14}$$

性质 3 $\hat{\beta}_L$ 的协方差矩阵是

$$\begin{aligned} \Sigma_{\hat{\beta}_L}\triangleq\operatorname{cov}(\hat{\beta}_L, \hat{\beta}_L)&=(X^\tau X)^{-1}X^\tau\operatorname{cov}(Y, Y)X(X^\tau X)^{-1} \\ &=(X^\tau X)^{-1}X^\tau\cdot\sigma^2I_nX(X^\tau X)^{-1} \\ &=\sigma^2(X^\tau X)^{-1} \end{aligned} \tag{15}$$

性质 4 $\hat{\beta}_L$ 是 β 的最小方差线性无偏估计．

在这里，若 T 是 β 的另一线性无偏估计，有

$$\Sigma_{\hat{\beta}_L}\leqslant\Sigma_T=\operatorname{cov}(T, T)$$

其中"$\leqslant$"表示 $\Sigma_T-\Sigma_{\hat{\beta}_L}$ 是非负定矩阵，并称 $\hat{\beta}_L$ 的方差不大于 T 的方差．

下面证明这一性质．

设 T 是 β 的任一线性无偏估计，则 T 必可表为

$$T=CY$$

其中 C 是 $m\times n$ 矩阵，由无偏性要求得

$$ET=E(CY)=CEY=CX\beta=\beta \tag{16}$$

由于(16)式要求对一切 β 成立，所以必须成立：

$$CX=I_m \tag{17}$$

而 $$\Sigma_T=\mathrm{cov}(T,T)=C\cdot\mathrm{cov}(Y,Y)\cdot C^\tau=\sigma^2(CC^\tau) \tag{18}$$

因为 $$\begin{aligned}0&\leqslant[C-(X^\tau X)^{-1}X^\tau][C-(X^\tau X)^{-1}X^\tau]^\tau\\&=(CC^\tau)+(X^\tau X)^{-1}-(X^\tau X)^{-1}X^\tau C^\tau-CX(X^\tau X)^{-1}\\&=(CC^\tau)+(X^\tau X)^{-1}-(X^\tau X)^{-1}-(X^\tau X)^{-1}\\&=(CC^\tau)-(X^\tau X)^{-1}\end{aligned}$$

所以

$$CC^\tau\geqslant(X^\tau X)^{-1} \tag{19}$$

联合(15)，(18)，(19)式，得到

$$\Sigma_{\hat\beta_L}=\sigma^2(X^\tau X)^{-1}\leqslant\sigma^2(CC^\tau)=\Sigma_T$$

由于 T 是任意选取的一个线性无偏估计，所以最小二乘估计 $\hat\beta_L$ 是 β 的最小方差线性无偏估计．这性质通常称为高斯-马尔科夫定理．

顺便指出，如果 Y 服从正态分布，即 $Y_i, i=1,\cdots,n$ 是相互独立的正态变量，具有相同的方差 σ^2，而 $EY_i=\sum_{j=1}^m x_{ij}\beta_j$，则由(12)式给出的 $\hat\beta_L$ 也是 β 的最小方差无偏估计．

下面给出最小二乘估计的几何解释

记 n 维向量 $x_j\triangleq\begin{pmatrix}x_{1j}\\ \vdots\\ x_{nj}\end{pmatrix}$，　　$j=1,\cdots,m$

它是矩阵 X 的第 j 列元素．用 $\mu(X)$ 表示由向量 x_j，$j=1,\cdots,m$ 的全部线性组合所构成的一个线性空间，则(9)式表示要求在 $\mu(X)$ 中寻找一个向量 $\xi=\sum_{j=1}^m\beta_jx_j$，使得 ξ 与 Y 之间的距离 $\|Y-$

$\xi\|$ 达到最小．而只有当 ξ 是 Y 在 $\mu(X)$ 中的投影时，才使 $\|Y-\xi\|$ 达到最小(见图 1)．

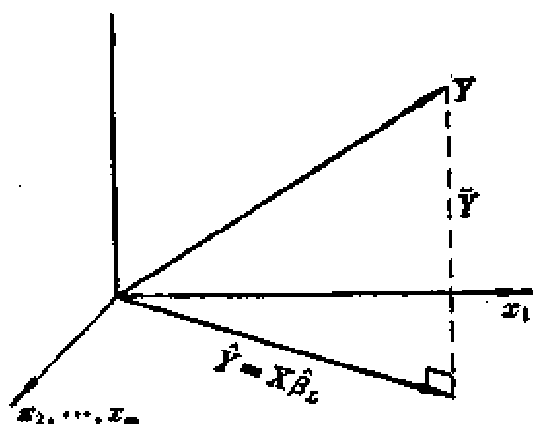


图 1

三、σ^2 的估计

记

$$\tilde{Y}\triangleq Y-X\hat{\beta}_L=[I_n-X(X^{\tau}X)^{-1}X^{\tau}]Y \tag{20}$$

$\tilde{Y}$ 称为剩余向量，它表示 Y 被估计后所剩下的残差，故又称为残差向量．又记 R_0^2 是剩余平方和

$$\begin{aligned}R_0^2\triangleq\tilde{Y}^{\tau}\tilde{Y}&=(Y-X\hat{\beta}_L)^{\tau}(Y-X\hat{\beta}_L)\\&=Y^{\tau}[I_n-X(X^{\tau}X)^{-1}X^{\tau}]Y\\&=Y^{\tau}Y-\hat{\beta}_L^{\tau}(X^{\tau}Y)\end{aligned} \tag{21}$$

对于 $\tilde{Y}$ 和 R_0^2，具有下列性质：

性质 1 $\tilde{Y}$ 和 $\hat{\beta}_L$ 互不相关．

$$\begin{aligned}\mathrm{cov}(\tilde{Y},\hat{\beta}_L)&=[I_n-X(X^{\tau}X)^{-1}X^{\tau}]\mathrm{cov}(Y,Y)[(X^{\tau}X)^{-1}X^{\tau}]^{\tau}\\&=\sigma^2[I_n-X(X^{\tau}X)^{-1}X^{\tau}][X(X^{\tau}X)^{-1}]=0\end{aligned} \tag{22}$$

从几何解释来看这一性质是显然的，那里 $\tilde{Y}$ 表示 Y 的垂线．(22)式意味着图 1 中的 $\hat{Y}$ 和 $\tilde{Y}$ 互相垂直．

性质 2 残差 $\tilde{Y}$ 的均值向量和协方差阵分别是

$$E(\tilde{Y})=E(Y-X\hat{\beta}_L)=0 \tag{23}$$

$$\text{cov}(\tilde{Y}, \tilde{Y}) = D[(I_n - X(X^\tau X)^{-1}X^\tau)Y]$$
$$= \sigma^2\{I_n - X(X^\tau X)^{-1}X^\tau\} \tag{24}$$

性质 3 记

$$\hat{\sigma}^2 \triangleq \frac{1}{n-m} R_0^2 \tag{25}$$

称为剩余方差，它是 σ^2 的无偏估计．即

$$E(\hat{\sigma}^2) = \sigma^2$$

因为
$$E(R_0^2) = E(\tilde{Y}^\tau \tilde{Y}) = \text{tr}\{\text{cov}(\tilde{Y}, \tilde{Y})\}$$
$$= \sigma^2 \text{tr}\{I_n - X(X^\tau X)^{-1}X^\tau\}$$
$$= \sigma^2(n - \text{tr} I_m) = \sigma^2(n-m)$$

这里记号 $\text{tr}\, A \triangleq \sum_{i=1}^{n} a_{ii}$ 称为 $n \times n$ 矩阵 A 的迹．

由此即得

$$E(\hat{\sigma}^2) = \frac{1}{n-m} E(R_0^2) = \sigma^2$$

它也可以从下面的定理 1 获得．

定理 1 对于由 (6) 式和 (7) 式给定的线性模型 $(Y, X\beta, \sigma^2 I_n)$，其中 X 是满列秩阵，且假定 Y 服从正态分布，则

(1) $\hat{\beta}_L$ 和 $\tilde{Y}$ 相互独立，同服从于正态分布．均值和协方差阵分别为

$$E\hat{\beta}_L = \beta, \quad E\tilde{Y} = 0 \tag{26}$$
$$\Sigma_{\hat{\beta}_L} = \sigma^2(X^\tau X)^{-1}$$
$$\text{cov}(\tilde{Y}, \tilde{Y}) = \sigma^2[I_n - X(X^\tau X)^{-1}X^\tau] \tag{27}$$

(2) $\hat{\beta}_L$ 和 R_0^2 相互独立；

(3) R_0^2/σ^2 服从于 $\chi^2(n-m)$ 分布．

[证明] (1)，(2) 显然，现证 (3)．

记
$$A \triangleq X(X^\tau X)^{-1}X^\tau$$

由于，A 是 $n\times n$ 非负定阵，秩为 m．所以存在 n 阶正交阵 C 使得

$$CAC^{\tau}=\begin{pmatrix}\lambda_1 & & & & & 0\\ & \ddots & & & & \\ & & \lambda_m & & & \\ & & & 0 & & \\ & & & & \ddots & \\ 0 & & & & & 0\end{pmatrix} \tag{28}$$

$$C^{\tau}C=I_n \tag{29}$$

其中 $\lambda_i>0$, $i=1\cdots m$.

而 $$A^2=AA=X(X^{\tau}X)^{-1}X^{\tau}\cdot X(X^{\tau}X)^{-1}X^{\tau}=A$$

对上式两边左乘 C 右乘 C^{τ} 得

$$CA^2C^{\tau}=CAC^{\tau}$$

而 $$CA^2C^{\tau}=CAC^{\tau}CAC^{\tau}=\begin{pmatrix}\lambda_1^2 & & & & & 0\\ & \ddots & & & & \\ & & \lambda_m^2 & & & \\ & & & 0 & & \\ & & & & \ddots & \\ 0 & & & & & 0\end{pmatrix}$$

$$CAC^{\tau}=\begin{pmatrix}\lambda_1 & & & & & 0\\ & \ddots & & & & \\ & & \lambda_m & & & \\ & & & 0 & & \\ & & & & \ddots & \\ 0 & & & & & 0\end{pmatrix}$$

所以 $$\lambda_i=\lambda_i^2,\quad i=1,\cdots,m$$

$$\lambda_i=1,\quad i=1,\cdots,m$$

因此(28)式实际上是

$$CAC^{\tau}=\begin{pmatrix}I_m & 0\\ 0 & 0\end{pmatrix} \tag{30}$$

作变量变换

$$Z \triangleq \begin{pmatrix} Z_1 \\ \vdots \\ Z_n \end{pmatrix} \triangleq C(Y - X\beta) \tag{31}$$

则

$$EZ = CE(Y - X\beta) = 0$$

$$\operatorname{cov}(Z, Z) = C \cdot \sigma^2 I_n C^\tau = \sigma^2 I_n$$

这表明 $Z_1, \cdots, Z_n$ 是相互独立，同服从于 $N(0, \sigma^2)$ 分布，且因 $Y = C^\tau Z + X\beta$，故

$$\begin{aligned} X\hat{\beta}_L - X\beta &= X(X^\tau X)^{-1}X^\tau Y - X\beta \\ &= X(X^\tau X)^{-1}X^\tau (Y - X\beta) \\ &= X(X^\tau X)^{-1}X^\tau C^\tau Z \end{aligned}$$

$$\begin{aligned} \|X\hat{\beta}_L - X\beta\|^2 &= (X\hat{\beta}_L - X\beta)^\tau (X\hat{\beta}_L - X\beta) \\ &= Z^\tau C X(X^\tau X)^{-1}X^\tau C^\tau Z \\ &= Z^\tau \begin{pmatrix} I_m & 0 \\ 0 & 0 \end{pmatrix} Z = Z_1^2 + \cdots + Z_m^2 \end{aligned}$$

又因 $\quad Y - X\beta = C^\tau Z$，及 $[I_n - X(X^\tau X)^{-1}X^\tau]X = 0$，

所以

$$\begin{aligned} R_0^2 &= Y^\tau [I_n - X(X^\tau X)^{-1}X^\tau] Y \\ &= (Y - X\beta)^\tau [I_n - X(X^\tau X)^{-1}X^\tau](Y - X\beta) \\ &= Z^\tau C[I_n - X(X^\tau X)^{-1}X^\tau]C^\tau Z \\ &= (Z_1^2 + \cdots + Z_n^2) - (Z_1^2 + \cdots + Z_m^2) \\ &= Z_{m+1}^2 + \cdots + Z_n^2 \end{aligned}$$

$$\|Y - X\beta\|^2 = \sum_{i=1}^n Z_i^2 = R_0^2 + \|X\hat{\beta}_L - X\beta\|^2 \tag{32}$$

所以 R_0^2/σ^2 服从于 $\chi^2(n-m)$ 分布.

$$E(R_0^2/\sigma^2) = n - m,$$

$$E(\hat{\sigma}^2) = E\left(\frac{1}{n-m} R_0^2\right) = \sigma^2$$

即剩余方差 $\hat{\sigma}^2$ 是方差 σ^2 的无偏估计.

四、参数带有约束的最小二乘估计

对于线性模型$(Y, X\beta, \sigma^2 I_n)$，在许多问题中，要求其中的未知参数 β 满足某特定的线性约束条件：$H^\tau\beta=r_0$. 这里 $H=(h_{ij})$ 是 $m\times s$ 矩阵 $(s<m)$，并假定它的秩为 s，$r_0^\tau=(r_{01}, \cdots, r_{0s})$ 是一个给定的 s 维向量. 常常希望求 β 的估计 $\hat{\beta}_H$，使得

$$\|Y-X\hat{\beta}_H\|^2=\min_{(\beta: H^\tau\beta=r_0)}\|Y-X\beta\|^2 \tag{33}$$

满足条件(33)的 $\hat{\beta}_H$ 称为 β 的具有线性约束 $H^\tau\beta=r_0$ 的最小二乘估计.

解 $\hat{\beta}_H$ 的问题实际上是在约束条件

$$H^\tau\beta=r_0 \tag{34}$$

下求
$$f=\|Y-X\beta\|^2=\sum_{i=1}^{n}\left(Y_i-\sum_{j=1}^{m}x_{ij}\beta_j\right)^2$$

的限制极值点问题. 用拉格朗日乘子法，记

$$\lambda\triangleq\begin{pmatrix}\lambda_1\\ \vdots\\ \lambda_s\end{pmatrix}, L\triangleq\sum_{i=1}^{n}\left(Y_i-\sum_{j=1}^{m}x_{ij}\beta_j\right)^2+2\sum_{l=1}^{s}\lambda_l\left(\sum_{j=1}^{m}h_{jl}\beta_j-r_{0l}\right)$$

令 L 关于 $\beta_1, \cdots, \beta_m$ 和 $\lambda_1, \cdots, \lambda_s$ 的偏导数为零，得方程组

$$\begin{cases}-\sum_{i=1}^{n}\left(Y_i-\sum_{j=1}^{m}x_{ij}\beta_j\right)x_{ik}+\sum_{l=1}^{s}\lambda_l h_{kl}=0, & k=1,\cdots,m\\ \sum_{j=1}^{m}h_{jl}\beta_j-r_{0l}=0, & l=1,\cdots,s\end{cases}$$

用向量矩阵形式表示，得

$$X^\tau X\beta+H\lambda=X^\tau Y \tag{35}$$

$$H^\tau\beta=r_0 \tag{36}$$

记 $\hat{\beta}_H$ 和 $\hat{\lambda}_H$ 是方程组 (35), (36) 的解，此时 $\hat{\beta}_H$ 就是在约束条件(34)之下，β 的最小二乘估计.

由(35)式得到

$$\hat{\beta}_H = (X^\tau X)^{-1}X^\tau Y - (X^\tau X)^{-1}H\hat{\lambda}_H$$
$$= \hat{\beta}_L - (X^\tau X)^{-1}H\hat{\lambda}_H \quad (37)$$

代入(36)式得

$$r_0 = H^\tau \hat{\beta}_H = H^\tau \hat{\beta}_L - H^\tau (X^\tau X)^{-1}H\hat{\lambda}_H$$

由于 H 是 $m\times s$ 满列秩阵，所以 $H^\tau(X^\tau X)^{-1}H$ 是 $s\times s$ 满秩阵，因此

$$\hat{\lambda}_H = [H^\tau(X^\tau X)^{-1}H]^{-1}(H^\tau\hat{\beta}_L - r_0) \quad (38)$$

代入(37)式得到

$$\hat{\beta}_H = \hat{\beta}_L - (X^\tau X)^{-1}H[H^\tau(X^\tau X)^{-1}H]^{-1}(H^\tau\hat{\beta}_L - r_0) \quad (39)$$

引理1 $\|Y-X\beta\|^2$ 可以分解为

$$\|Y-X\beta\|^2 = \|Y-X\hat{\beta}_L\|^2 + \|X(\hat{\beta}_L-\hat{\beta}_H)\|^2 + \|X(\hat{\beta}_H-\beta)\|^2 \quad (40)$$

［证明］ 由于

$$\|X(\hat{\beta}_L-\beta)\|^2 = (\hat{\beta}_L-\beta)^\tau X^\tau X(\hat{\beta}_L-\beta)$$
$$= (\hat{\beta}_L-\hat{\beta}_H+\hat{\beta}_H-\beta)^\tau X^\tau X(\hat{\beta}_L-\hat{\beta}_H+\hat{\beta}_H-\beta)$$
$$= (\hat{\beta}_L-\hat{\beta}_H)^\tau X^\tau X(\hat{\beta}_L-\hat{\beta}_H)$$
$$+ (\hat{\beta}_H-\beta)X^\tau X(\hat{\beta}_H-\beta)$$
$$+2(\hat{\beta}_L-\hat{\beta}_H)^\tau X^\tau X(\hat{\beta}_H-\beta)$$
$$= \|X(\hat{\beta}_L-\hat{\beta}_H)\|^2 + \|X(\hat{\beta}_H-\beta)\|^2$$
$$+2(\hat{\beta}_L-\hat{\beta}_H)^\tau X^\tau X(\hat{\beta}_H-\beta)$$

而由(37)，(38)式得

$$(\hat{\beta}_L-\hat{\beta}_H)^\tau X^\tau X(\hat{\beta}_H-\beta) = \hat{\lambda}_H^\tau H^\tau(\hat{\beta}_H-\beta) = \hat{\lambda}_H^\tau(r_0-r_0) = 0$$

所以 $\|X(\hat{\beta}_L-\beta)\|^2 = \|X(\hat{\beta}_L-\hat{\beta}_H)\|^2 + \|X(\hat{\beta}_H-\beta)\|^2$

将上式代入(32)式立即可得(40)式成立.

由(40)式可以看到，当 $\beta=\hat{\beta}_H$ 时，有

$$R_1^2 \triangleq \min_{(\beta:H^\tau\beta=r_0)} \|Y-X\beta\|^2 = \|Y-X\hat{\beta}_H\|^2$$
$$= R_0^2 + \|X(\hat{\beta}_L-\hat{\beta}_H)\|^2 \quad (41)$$

下面计算 $E(R_1^2)$，并获得在约束条件成立的条件下，σ^2 的无偏估计，利用(39)式

$$\begin{aligned}
\|X(\hat{\beta}_L-\hat{\beta}_H)\|^2 &= (H^{\tau}\hat{\beta}_L-r_0)^{\tau}[H^{\tau}(X^{\tau}X)^{-1}H]^{-1}(H^{\tau}\hat{\beta}_L-r_0)\\
&=(H^{\tau}\hat{\beta}_L-H^{\tau}\beta+H^{\tau}\beta-r_0)^{\tau}[H^{\tau}(X^{\tau}X)^{-1}H]^{-1}\\
&\quad\times(H^{\tau}\hat{\beta}_L-H^{\tau}\beta+H^{\tau}\beta-r_0)\\
&=[H^{\tau}(\hat{\beta}_L-\beta)]^{\tau}[H^{\tau}(X^{\tau}X)^{-1}H]^{-1}[H^{\tau}(\hat{\beta}_L-\beta)]\\
&\quad+(H^{\tau}\beta-r_0)^{\tau}[H^{\tau}(X^{\tau}X)^{-1}H]^{-1}(H^{\tau}\beta-r_0)\\
&\quad+2[H^{\tau}(\hat{\beta}_L-\beta)]^{\tau}[H^{\tau}(X^{\tau}X)^{-1}H]^{-1}(H^{\tau}\beta-r_0)
\end{aligned}$$

由(14)，(15)式得到

$$E\{[H^{\tau}(\hat{\beta}_L-\beta)]^{\tau}[H^{\tau}(X^{\tau}X)^{-1}H]^{-1}(H^{\tau}\beta-r_0)\}=0$$

$$\begin{aligned}
&E\{[H^{\tau}(\hat{\beta}_L-\beta)]^{\tau}[H^{\tau}(X^{\tau}X)^{-1}H]^{-1}[H^{\tau}(\hat{\beta}_L-\beta)]\}\\
&=E\{(\hat{\beta}_L-\beta)^{\tau}H[H^{\tau}(X^{\tau}X)^{-1}H]^{-1}H^{\tau}(\hat{\beta}_L-\beta)\}\\
&=\mathrm{tr}\{H[H^{\tau}(X^{\tau}X)^{-1}H]^{-1}H^{\tau}\Sigma_{\hat{\beta}_L}\}\\
&=\mathrm{tr}\{H[H^{\tau}(X^{\tau}X)^{-1}H]^{-1}H^{\tau}(X^{\tau}X)^{-1}\sigma^2\}\\
&=\sigma^2\mathrm{tr}\{[H^{\tau}(X^{\tau}X)^{-1}H]^{-1}H^{\tau}(X^{\tau}X)^{-1}H\}\\
&=\sigma^2\mathrm{tr}I_s=s\cdot\sigma^2
\end{aligned}$$

因此

$$\begin{aligned}
E\{\|X(\hat{\beta}_L-\hat{\beta}_H)\|^2\}&=s\cdot\sigma^2\\
&\quad+(H^{\tau}\beta-r_0)^{\tau}[H^{\tau}(X^{\tau}X)^{-1}H]^{-1}(H^{\tau}\beta-r_0)
\end{aligned}$$

$$\begin{aligned}
E(R_1^2)&=E(R_0^2)+E\{\|X(\hat{\beta}_L-\hat{\beta}_H)\|^2\}\\
&=(n-m+s)\sigma^2+(H^{\tau}\beta-r_0)^{\tau}[H^{\tau}(X^{\tau}X)^{-1}H]^{-1}\\
&\quad\times(H^{\tau}\beta-r_0)
\end{aligned}$$

所以当约束条件(34)成立时，

$$E(R_1^2)=(n-m+s)\sigma^2$$

因此

$$\boxed{\hat{\sigma}_1^2=\frac{1}{n-m+s}R_1^2}$$

是 σ^2 的无偏估计。

类似于定理 1，有

定理 2 关于线性模型 $(Y, X\beta, \sigma^2 I_n)$，其中 X 是 $n\times m$ ($n>m$) 满列秩阵，Y 是正态变量. 记约束条件是 $H^\tau\beta=r_0$，其中 H 是 $m\times s(m>s)$ 满列秩阵，则对于由(21)式和(41)式决定的 R_0^2 和 R_1^2，有

(1) $R_1^2-R_0^2=(H^\tau\hat{\beta}_L-r_0)^\tau[H^\tau(X^\tau X)^{-1}H]^{-1}(H^\tau\hat{\beta}_L-r_0)$,
$E(R_1^2-R_0^2)=\sigma^2 s+(H^\tau\beta-r_0)^\tau[H^\tau(X^\tau X)^{-1}H]^{-1}(H^\tau\beta-r_0)$

(2) 当约束条件 $H^\tau\beta=r_0$ 为真时，$R_1^2-R_0^2$ 与 R_0^2 相互独立，$(R_1^2-R_0^2)/\sigma^2$ 服从于 $\chi^2(s)$ 分布，R_1^2/σ^2 服从于 $\chi^2(n-m+s)$ 分布，

$$F=\frac{R_1^2-R_0^2}{R_0^2}\cdot\frac{n-m}{s} \tag{42}$$

服从于 $F(s, n-m)$ 分布.

[证明] (1)上面已证明，现证(2).

由定理 1，$\hat{\beta}_L$ 服从 $N(\beta, (X^\tau X)^{-1}\sigma^2)$分布，因此

$$\begin{aligned}\operatorname{cov}(H^\tau\hat{\beta}_L, H^\tau\hat{\beta}_L)&=H^\tau\operatorname{cov}(\hat{\beta}_L, \hat{\beta}_L)H\\&=H^\tau(X^\tau X)^{-1}H\sigma^2\end{aligned}$$

所以当 $H^\tau\beta=r_0$ 为真时，$H^\tau\hat{\beta}_L$ 服从 $N(r_0, H^\tau(X^\tau X)^{-1}H\sigma^2)$分布，由第 2 章习题 5 知道

$$(H^\tau\hat{\beta}_L-r_0)^\tau\{H^\tau(X^\tau X)^{-1}H\sigma^2\}^{-1}(H^\tau\hat{\beta}_L-r_0)=\frac{R_1^2-R_0^2}{\sigma^2} \tag{43}$$

服从于自由度为 s 的 χ^2-分布.

由于 $R_1^2-R_0^2$ 是 $\hat{\beta}_L$ 的函数，所以由定理 1 知道，它与 R_0^2 独立，R_0^2/σ^2 服从 $\chi^2(n-m)$ 分布，因此由 (42) 式决定的统计量 F 服从 $F(s, n-m)$ 分布. 并由定理 2.2.3 知道，R_1^2/σ^2 服从于 $\chi^2(n-m+s)$分布.

五、关于参数 β 的假设检验和置信区间

下面我们假定 Y 服从正态分布

1．一个参数函数的推断．

设 $\theta=\sum_{i=1}^{m}a_i\beta_i$ 是未知参数 β 的某一给定的线性函数，θ 的最小二乘估计是

$$\hat{\theta}_L=(a_1,\ \cdots,\ a_m)\hat{\beta}_L \tag{44}$$

显然它是 θ 的线性无偏估计，其方差为

$$D(\hat{\theta}_L)=\sigma^2(a_1,\ \cdots,\ a_m)(X^\tau X)^{-1}\begin{pmatrix}a_1\\ \vdots\\ a_m\end{pmatrix}\triangleq\sigma^2d^2$$

由定理 1 知道

$$T=\frac{\hat{\theta}_L-\theta}{\sigma d}\Bigg/\sqrt{\frac{R_0^2}{\sigma^2(n-m)}}=\frac{\hat{\theta}_L-\theta}{R_0d}\sqrt{n-m}\sim t(n-m) \tag{45}$$

利用 §3.2 和 §3.3 的结果，有

关于检验假设 $H_0(\theta=\theta_0)$，则我们当 $\frac{|\hat{\theta}_L-\theta_0|}{R_0d}\sqrt{n-m}>t_{\alpha/2}$ 时拒绝假设 $H_0(\theta=\theta_0)$，否则认为试验结果与假设 H_0 无显著差异．

而 θ 的置信度为 $1-\alpha$ 的置信区间是 $(\hat{\theta}_1,\ \hat{\theta}_2)$，

其中
$$\hat{\theta}_1=\hat{\theta}_L-t_{\alpha/2}\cdot\frac{R_0d}{\sqrt{n-m}}$$
$$\hat{\theta}_2=\hat{\theta}_L+t_{\alpha/2}\cdot\frac{R_0d}{\sqrt{n-m}}$$

特别地取 $\theta=\beta_i$，即考虑假设 $H_0(\beta_i=\beta_{i0})$，由 (45) 式得到 T 统计量是

$$T=\frac{\hat{\beta}_i-\beta_{i0}}{R_0\sqrt{C_{ii}}}\sqrt{n-m} \tag{46}$$

其中 C_{ii} 是方阵 $C\triangleq(X^{\tau}X)^{-1}$ 对角线上第 i 个元素.

类似地，β_i 的置信区间是

$$\left(\hat{\beta}_i-t_{\alpha/2}\cdot\frac{R_0\sqrt{C_{ii}}}{\sqrt{n-m}},\ \hat{\beta}_i+t_{\alpha/2}\frac{R_0\sqrt{C_{ii}}}{\sqrt{n-m}}\right)$$

2．几个参数函数的推断

设 $\theta=H^{\tau}\beta$, H 是已知的 $m\times s(s<m)$ 矩阵，且秩为 s, $\theta^{\tau}=(\theta_1,\cdots,\theta_s)$, $\theta_1,\cdots,\theta_s$ 是 β 的 s 个确定的线性函数，它的最小二乘估计是

$$\hat{\theta}_L=H^{\tau}\hat{\beta}_L \tag{47}$$

$$\begin{gathered} E\hat{\theta}_L=H^{\tau}E(\hat{\beta}_L)=H^{\tau}\beta=\theta \\ D(\hat{\theta}_L)=H^{\tau}D(\hat{\beta}_L)H=\sigma^2H^{\tau}(X^{\tau}X)^{-1}H\triangleq\sigma^2D_{\theta_L} \end{gathered} \tag{48}$$

运用第 2 章习题 5, $(\hat{\theta}_L-\theta)^{\tau}D_{\theta_L}^{-1}(\hat{\theta}_L-\theta)/\sigma^2$ 服从 $\chi^2(s)$分布，且由定理 1 知道它与 R_0^2 独立，所以

$$F=\frac{(\hat{\theta}_L-\theta)^{\tau}D_{\theta_L}^{-1}(\hat{\theta}_L-\theta)}{R_0^2}\ \frac{n-m}{s} \tag{49}$$

服从 $F(s,n-m)$ 分布.

关于假设 H_0: $\theta=\theta_0$.

考察偏差 $\hat{\theta}_L-\theta_0$, 则运用(48)式，容易计算

$$E\left\{\frac{1}{s}(\hat{\theta}_L-\theta_0)^{\tau}D_{\theta_L}^{-1}(\hat{\theta}_L-\theta)\right\}$$

$$=\sigma^2+\frac{1}{s}(H^{\tau}\beta-\theta_0)^{\tau}D_{\theta_L}^{-1}(H^{\tau}\beta-\theta_0)$$

$$=\begin{cases}\sigma^2, & \text{当 } H_0 \text{ 为真时}\\ \sigma^2+\dfrac{1}{s}(H^{\tau}\beta-\theta_0)^{\tau}D_{\theta_L}^{-1}(H^{\tau}\beta-\theta_0)>\sigma^2, & \text{其它}\end{cases}$$

而 $E\left\{\dfrac{1}{n-m}R_0^2\right\}=\sigma^2$, 与假设 H_0 无关．所以 F 统计量 (49) 在假设 H_0 不真时有偏大的趋势，也就是说，F 统计量的大值表明试验结果与假设 H_0 有偏离．因此对水平为 α, 当 $F\geqslant F_\alpha$ 时拒绝假

设 H_0.

利用定理 2, 也可采用下面的方差分析表①

表 1

方差来源	平方和	自由度	平均平方和
假设 $H^\tau\beta=\theta_0$ 引起的离差	$R_1^2-R_0^2$	s	$\frac{1}{s}(R_1^2-R_0^2)$
剩余	$R_0^2=\min\limits_{\beta}(Y-X\beta)^\tau(Y-X\beta)$	$n-m$	$\frac{1}{n-m}R_0^2$
总和	$R_1^2=\min\limits_{\{\beta:H^\tau\beta=\theta_0\}}(Y-X\beta)^\tau\times(Y-X\beta)$	$n-m+s$	

则当假设 $H_0(\theta=\theta_0)$ 为真时, $(R_1^2-R_0^2)/\sigma^2$ 服从于 $\chi^2(s)$ 分布, 且与 R_0^2 独立, 所以统计量

$$F=\frac{R_1^2-R_0^2}{R_0^2}\cdot\frac{n-m}{s} \tag{50}$$

当 H_0 为真时, 它服从 $F(s,n-m)$ 分布, 与统计量 (49) 一样可作为检验假设 $H_0(H^\tau\beta=\theta_0)$ 的统计量. 实际上, 由(43)式可见

$$R_1^2-R_0^2=(\hat{\theta}_L-\theta_0)^\tau D_{\hat{\theta}_L}^{-1}(\hat{\theta}_L-\theta_0)$$

所以由(49)式给定的统计量与由(50)式给定的统计量是相同的.

特别地, 如果 $H^\tau\beta=\beta_i$, 欲检验的假设是 H_0: $\beta_i=0$, 则有

$$R_1^2-R_0^2=\min_{\beta_i=0}(Y-X\beta)^\tau(Y-X\beta)$$

$$-\min_{\beta}(Y-X\beta)^\tau(Y-X\beta)=\frac{1}{C_{ii}}\beta_i^2 \tag{51}$$

这个量又称为偏平方和, 它表示消除了 $X_1,\cdots,X_{i-1},X_{i+1},\cdots,X_m$ 的影响以后, X_i 对 Y 的影响. 其中 C_{ii} 同(46)式中的定义.

关于检验假设 $H_0(\beta_i=0)$, 由(50)式定义的检验统计量 F 可以写为

① 关于方差分析方法的详细讨论见 §3.

$$F=\frac{\hat{\beta}_i^2}{R_0^2 C_{ii}}(n-m) \tag{52}$$

当 $H_0(\beta_i=0)$ 为真时，它服从 $F(1, n-m)$ 分布.

六、关于 Y 的预测

仍以例 1 为例，在有些时候，我们需要通过过去不同时刻 t_1, …, t_n 对质点位置的观察数据 y_1, …, y_n，预测未来某 t_0 时刻质点的位置 S_0. 显然 S_0 的预测量可用相应的经验回归方程(13)式

$$\hat{S}_0=\hat{\alpha}_L+\hat{\beta}_L t_0$$

得到，其中 $\hat{\alpha}_L$ 和 $\hat{\beta}_L$ 是运用历史数据 (t_1, y_1), …, (t_n, y_n)所获得的 α 和 β 的最小二乘估计.

一般地，对于一组指定的 $x_0=(x_{01}, \cdots, x_{0m})^T$，由经验回归方程式(13)式，其对应的 Y_0 的预测值是

$$\hat{Y}_0=x_0^T\hat{\beta}_L=\sum_{j=1}^{m}x_{0j}\hat{\beta}_j \tag{53}$$

这个预测量的好坏依赖于它可能误差的大小，因而我们常常希望考虑这个误差而给出一个类似于置信区间的预测区间. 我们知道 Y_0 是随机变量，且假定它服从正态分布，它的分布是 $N(x_0^T\beta, \sigma^2)$. 而现在的预测量 $\hat{Y}_0$ 具有两个误差源：一个是来自于 Y_0 的均值估计 $x_0^T\hat{\beta}_L$，另一个是 Y_0 的实际值与它的均值之间也有偏离. 而均值估计的误差本质上是由 $\hat{\beta}_L$ 的抽样误差所决定. 如果 β 和 σ^2 是已知的，则 Y_0 的 95% 概率的预测区间是

$$(x_0^T\beta-1.96\sigma,\ x_0^T\beta+1.96\sigma)$$

此时 Y_0 属于这个区间的概率正好是 0.95. 而现在除了 x_0 以外，β 和 σ^2 都是未知参数. 因此就必需借助于它的估计来建立这样的区间.

令

$$Z \triangleq Y_0-x_0^T\hat{\beta}_L \tag{54}$$

则 Z 服从正态分布，且由于 Y_0 与 Y_1, …, Y_n 独立，$\hat{\beta}_L$ 与 R_0^2 独

立，所以 Z 与 R_0^2 独立，因此

$$EZ=EY_0-E(x_0^\tau\hat{\beta}_L)=x_0^\tau\beta-x_0^\tau\beta=0$$

$$\begin{aligned}D(Z)&=D(Y_0-x_0^\tau\hat{\beta}_L)=DY_0+D(x_0^\tau\hat{\beta}_L)\\&=\sigma^2+\sigma^2x_0^\tau(X^\tau X)^{-1}x_0\\&=\sigma^2\{1+x_0^\tau(X^\tau X)^{-1}x_0\}\end{aligned}\tag{55}$$

统计量

$$\begin{aligned}T&=\frac{Z}{\{1+x_0^\tau(X^\tau X)^{-1}x_0\}^{\frac{1}{2}}}\cdot\frac{1}{R_0/\sqrt{n-m}}\\&=\frac{Y_0-x_0^\tau\hat{\beta}_L}{[1+x_0^\tau(X^\tau X)^{-1}x_0]^{\frac{1}{2}}}\cdot\frac{1}{R_0/\sqrt{n-m}}\end{aligned}\tag{56}$$

服从于自由度为$(n-m)$的 t-分布．由此可以得到水平为 $1-\alpha$ 的预测区间：

$$(\hat{Y}_{01},\ \hat{Y}_{02})$$

其中

$$\left.\begin{aligned}\hat{Y}_{01}&=x_0^\tau\hat{\beta}_L-t_{\alpha/2}\cdot\frac{R_0}{\sqrt{n-m}}\cdot\{1+x_0^\tau(X^\tau X)^{-1}x_0\}^{1/2}\\\hat{Y}_{02}&=x_0^\tau\hat{\beta}_L+t_{\alpha/2}\cdot\frac{R_0}{\sqrt{n-m}}\{1+x_0^\tau(X^\tau X)^{-1}x_0\}^{1/2}\end{aligned}\right\}\tag{57}$$

它表示

$$P\{\hat{Y}_{01}<Y_0<\hat{Y}_{02}\}=1-\alpha\tag{58}$$

七、例子

[例 3][①]　在平炉炼钢过程中，由于矿石及炉气的氧化作用，铁水的总含碳量在不断降低．一炉钢在冶炼初期（溶化期）中总的去碳量 Y 与所加的两种矿石（天然矿石和烧结矿石）的加入量 X_1, X_2 及熔化时间 X_3（熔化时间愈长则炉气去碳量愈多）有关．经实测某号平炉的相应数据如下：

① 本例选自形季，"回归分析简介"，数学的实践与认识，第三期，1971 年．

表 2　某平炉冶炼初期总去碳量与矿石加入量及熔化时间的记录

编号	Y (吨)	X_1 (槽)	X_2 (槽)	X_3 (5分钟)	编号	Y (吨)	X_1 (槽)	X_2 (槽)	X_3 (5分钟)
1	4.3302	2	18	50	26	2.7066	9	6	39
2	3.6458	7	9	40	27	5.6314	12	5	51
3	4.4830	5	14	46	28	5.8152	6	13	41
4	5.5468	12	3	43	29	5.1302	12	7	47
5	5.4970	1	20	64	30	5.3910	0	24	61
6	3.1125	3	12	40	31	4.4583	5	12	37
7	5.1182	3	17	64	32	4.6569	4	15	49
8	3.8759	6	5	39	33	4.5212	0	20	45
9	4.6700	7	8	37	34	4.8650	6	16	42
10	4.9536	0	23	55	35	5.3566	4	17	48
11	5.0060	3	16	60	36	4.6098	10	4	48
12	5.2701	0	18	49	37	2.3815	4	14	36
13	5.3772	8	4	50	38	3.8746	5	13	36
14	5.4849	6	14	51	39	4.5919	9	18	51
15	4.5960	0	21	51	40	5.1588	6	13	54
16	5.6645	3	14	51	41	5.4373	5	18	100
17	6.0795	7	12	56	42	3.9960	5	11	44
18	3.2194	16	0	48	43	4.3970	8	6	63
19	5.8076	6	16	45	44	4.0622	2	13	55
20	4.7306	0	15	52	45	2.2905	7	8	50
21	4.6805	9	0	40	46	4.7115	4	10	45
22	3.1272	4	6	32	47	4.5310	10	5	40
23	2.6104	0	17	47	48	5.3637	3	17	64
24	3.7174	9	0	44	49	6.0771	4	15	72
25	3.8946	2	6	39					

由经验知道，Y与X_1, X_2, X_3之间可用下面这一线性关系描述

$$Y=\beta_0+\beta_1X_1+\beta_2X_2+\beta_3X_3+e \tag{59}$$

其中e是服从$N(0, \sigma^2)$分布的随机变量．

要求通过上表给出的实测记录，定出未知参数$\beta^{\tau}\triangleq(\beta_0, \beta_1, \beta_2, \beta_3)$和$\sigma^2$．

运用最小二乘法，正规方程是

$$\begin{pmatrix} n & \Sigma x_{i1} & \Sigma x_{i2} & \Sigma x_{i3} \\ \Sigma x_{i1} & \Sigma x_{i1}^2 & \Sigma x_{i1}x_{i2} & \Sigma x_{i1}x_{i3} \\ \Sigma x_{i2} & \Sigma x_{i2}x_{i1} & \Sigma x_{i2}^2 & \Sigma x_{i2}x_{i3} \\ \Sigma x_{i3} & \Sigma x_{i3}x_{i1} & \Sigma x_{i3}x_{i2} & \Sigma x_{i3}^2 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} = \begin{pmatrix} \Sigma Y_i \\ \Sigma Y_i x_{i1} \\ \Sigma Y_i x_{i2} \\ \Sigma Y_i x_{i3} \end{pmatrix} \tag{60}$$

由于关系式 (59) 中存在有常数项 β_0，所以在具体计算时，常常把(59)式改写为

$$Y_i = \mu + \beta_1(x_{i1} - \bar{x}_1) + \beta_2(x_{i2} - \bar{x}_2) + \beta_3(x_{i3} - \bar{x}_3) + e_i \tag{61}$$

其中

$$\bar{x}_j \triangleq \frac{1}{n}\sum_{i=1}^{n} x_{ij}, \ j = 1, 2, 3$$

$$\mu \triangleq \beta_0 + \beta_1\bar{x}_1 + \beta_2\bar{x}_2 + \beta_3\bar{x}_3$$

对应于(61)式，正规方程可改写为

$$\begin{pmatrix} n & 0 & 0 & 0 \\ 0 & l_{11} & l_{12} & l_{13} \\ 0 & l_{12} & l_{22} & l_{23} \\ 0 & l_{13} & l_{23} & l_{33} \end{pmatrix} \begin{pmatrix} \mu \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} = \begin{pmatrix} \Sigma Y_i \\ l_{01} \\ l_{02} \\ l_{03} \end{pmatrix} \tag{62}$$

其中

$$\bar{Y} \triangleq \frac{1}{n}\sum_{i=1}^{n} Y_i, \quad l_{00} \triangleq \sum_{i=1}^{n}(Y_i - \bar{Y})^2$$

$$l_{0j} \triangleq \sum_{i=1}^{n}(Y_i - \bar{Y})(x_{ij} - \bar{x}_j), \quad j = 1, 2, 3$$

$$l_{jk} \triangleq \sum_{i=1}^{n}(x_{ij} - \bar{x}_j)(x_{ik} - \bar{x}_k), \ j, k = 1, 2, 3$$

显然，方程(62)又可写为

$$\begin{pmatrix} l_{11} & l_{12} & l_{13} \\ l_{12} & l_{22} & l_{23} \\ l_{13} & l_{23} & l_{33} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} = \begin{pmatrix} l_{01} \\ l_{02} \\ l_{03} \end{pmatrix} \tag{63}$$

和

$$\beta_0 = \bar{Y} - \sum_{j=1}^{3} \beta_j \bar{x}_j \tag{64}$$

记
$$L \triangleq \begin{pmatrix} l_{11} & l_{12} & l_{13} \\ l_{12} & l_{22} & l_{23} \\ l_{13} & l_{23} & l_{33} \end{pmatrix}$$

则有

$$\begin{pmatrix} n & 0 \\ 0 & L \end{pmatrix}^{-1} = \begin{pmatrix} \frac{1}{n} & 0 \\ 0 & L^{-1} \end{pmatrix} \tag{65}$$

和
$$\begin{aligned} R_0^2 &= \sum_{i=1}^{n}[Y_i-(\hat\beta_0+\hat\beta_1x_{i1}+\hat\beta_2x_{i2}+\hat\beta_3x_{i3})]^2 \\ &= \sum_{i=1}^{n}\left[(Y_i-\overline{Y})-\sum_{j=1}^{3}\hat\beta_j(x_{ij}-\bar{x}_j)\right]^2 \\ &= l_{00}-\sum_{j=1}^{3}\hat\beta_j l_{0j} \end{aligned} \tag{66}$$

由表中数据计算得到

$$n=49,\ \Sigma Y_i=224.5169,\ \overline{Y}=4.582$$
$$\Sigma x_{i1}=259,\quad \bar{x}_1=5.286$$
$$\Sigma x_{i2}=578,\quad \bar{x}_2=11.796$$
$$\Sigma x_{i3}=2411,\ \bar{x}_3=49.204$$

$$l_{00}=44.905$$
$$l_{11}=662.00,\quad l_{12}=-918.1428,\quad l_{13}=-388.8571$$
$$l_{22}=1753.959,\quad l_{23}=776.0408,\quad l_{33}=6247.959$$
$$l_{01}=-6.432896,\ l_{02}=69.13047,\quad l_{03}=245.5713$$

代入方程组(63)得

$$\begin{pmatrix} 662.00 & -918.1428 & -388.8571 \\ -918.1428 & 1753.959 & 776.0408 \\ -388.8571 & 776.0408 & 6247.959 \end{pmatrix}\begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix}=\begin{pmatrix} -6.432986 \\ 69.13047 \\ 245.5713 \end{pmatrix}$$

和
$$\beta_0=4.582-5.286\beta_1-11.796\beta_2-49.204\beta_3$$

记系数阵
$$L \triangleq \begin{pmatrix} 662.00 & -918.1428 & -388.8571 \\ -918.1428 & 1753.959 & 776.0408 \\ -388.8571 & 776.0408 & 6247.959 \end{pmatrix}$$

则容易算得 L 的逆阵是

$$C=(C_{ij})\triangleq L^{-1}=\begin{pmatrix} 0.005515 & 0.002894 & -0.00001623 \\ 0.002894 & 0.002122 & -0.00008345 \\ -0.00001623 & -0.00008345 & 0.0001694 \end{pmatrix}$$

β 的最小二乘估计值是

$$\hat{\beta}_1=0.1606,\quad \hat{\beta}_2=0.1076$$

$$\hat{\beta}_3=0.0359,\quad \hat{\beta}_0=0.7014$$

所以 Y 关于 X_1, X_2, X_3 的经验关系式是

$$\hat{Y}=0.7014+0.1606X_1+0.1076X_2+0.0359X_3$$

由(65)式得

$$R_0^2=l_{00}-\sum_{j=1}^{3}l_{0j}\hat{\beta}_j=44.905-15.230=29.675$$

$$\hat{\sigma}^2=\frac{R_0^2}{n-m}=\frac{1}{45}29.675=0.659$$

$$\hat{\sigma}=0.81$$

下面考虑检验问题，例如要求检验两种矿石加入量的大小对初期总去碳量有无显著影响，这个问题就归结为检验假设

$$H_0:\ \begin{pmatrix}\beta_1\\ \beta_2\end{pmatrix}=0$$

此时，

$$R_1^2=\min_{\beta_1=0,\beta_2=0}(Y-X\beta)^{\tau}(Y-X\beta)$$

$$=\min_{\beta_0\beta_3}\sum_{i=1}^{49}[Y_i-(\beta_0+\beta_3x_{i3})]^2$$

$$=l_{00}-\hat{\beta}_3'l_{03}$$

其中

$$\hat{\beta}_3'=\frac{l_{03}}{l_{33}}=\frac{245.5713}{6247.959}=0.0393$$

$$R_1^2=44.905-9.651=35.254$$

$$R_1^2-R_0^2=35.254-29.675=5.579$$

代入(50)式得

$$F=\frac{R_1^2-R_0^2}{R_0^2}\cdot\frac{45}{2}=4.23$$

对水平 $\alpha=0.05$，自由度为 (2, 45)，查附表 4 得 $F_\alpha=3.205$. 现在 $F=4.23>F_\alpha$，故拒绝假设 $H_0\left(\begin{pmatrix}\beta_1\\ \beta_2\end{pmatrix}=0\right)$，这表示两种矿石加入量的多少对冶炼初期总去碳量具有显著影响.

在结束这一节时，还想指出这节所处理的模型的特点是要求关于未知参数呈线性关系. 也就是说如果因变量 Y 和自变量 X_1, $\cdots$, X_m 之间服从关系式

$$Y=f(X_1,\ \cdots,\ X_m,\ \beta_0,\ \beta_1,\ \cdots,\ \beta_m)+e$$

则要求 f 是未知参数 β_0, $\cdots$, β_m 的线性函数. 在有些应用中 f 虽不是 $\beta_0,\cdots,\beta_m$ 的线性函数,但是通过某种变换可以化为线性函数,也仍然可以用这节所介绍的方法处理. 例如要研究猪的体重 (M) 与它的身长(L)和腰围 (R) 之间的关系. 从经验知道它们之间应具有关系式

$$M=C\cdot L^{\beta_1}\cdot R^{\beta_2}+e$$

其中 $f(C,\ \beta_1,\ \beta_2)=CL^{\beta_1}R^{\beta_2}$，它不是未知参数 C, β_1, β_2 的线性函数,因此它不能用这节的方法处理. 但只要对它取对数,就有

$$\log M=\log C+\beta_1\log L+\beta_2\log R+e'$$

记 $Y\triangleq\log M$, $\beta_0\triangleq\log C$, $X_1\triangleq\log L$, $X_2\triangleq\log R$, 则上式又可写为

$$Y=\beta_0+\beta_1X_1+\beta_2X_2+e'$$

它是个线性模型,因此关于 β_0, β_1, β_2 可以获得最小二乘估计 $\hat\beta_0$, $\hat\beta_1$, $\hat\beta_2$, 相应地就可以得到未知参数的估计 $\hat C=e^{\hat\beta_0}$, $\hat\beta_1$, $\hat\beta_2$.

在利用最小二乘法拟合数学模型时，有时可能有很多因子对因变量 Y 起作用. 这时就要求从众多的有可能影响 Y 的诸因子中挑选出一批对 Y 起主要作用的因子，以建立起一个优良的数学模型. 对此可利用逐步回归方法处理①.

① 可参看上海师范大学数学系概率统计教研组,《回归分析及其试验设计》.

§2. 回 归 分 析

一、回归的概念

在上节中，讨论了线性模型(1.4)～(1.7)的统计推断问题.在线性模型中，我们假定自变量 X_i 是可以测量和控制的非随机变量，Y 是随机变量，且已知 Y 和 X_i 之间服从某种线性关系. 要求通过 X_i 和 Y 的观察，估计其中的未知参数 β 或检验某 X_{i0} 对 Y 是否有影响；对某些特定的 X_i 的值预测 Y 等等. 在许多实际问题中，还经常要求处理另一类十分相似的统计模型，所不同的是在那里自变量 X 也是随机变量；类似地要求研究估计、检验、预测等统计问题. 例如：

在气象学中，希望基于在某几个时刻对大气压的测量，要求预测几小时以后的气象情况；

在纺织工业中，大家知道纱的质量是与原棉的质量具有很密切的关系，而原棉的质量往往与其品种、产地有关. 因此为了预测和控制纱的质量，就必需研究棉花的质量指标(如棉花成熟度、细度、强力、长度均匀度、杂质等)与纺出的纱的品质指标之间的关系，从而通过对各种原棉的合理搭配，以保证所纺出的纱合乎要求.

在以上二例中，都要求研究某个随机变量 Y(如几小时后的气象情况或纱的质量)与另一些随机变量 $X_1, \cdots, X_m$ (如大气压或原棉的成熟度、细度、强力、长度均匀度、杂质)之间的相互关系，并通过对 $X_1, \cdots, X_m$ 的测量，预测在今后可以观察的变量 Y. 在这里，一个自然的问题是如何选择一个好的预测器呢？

设随机变量 X, Y, $X^\tau=(X_1, \cdots, X_m)$是 m 维随机向量，它是可以预先测量的，希望通过 X 预测 Y，也就是说要寻找一个函数 $y=M(x)=M(x_1, \cdots, x_m)$，当 X 的观察值为 x 时，就把 $M(x)$ 作为对 Y 的预测值. 当然一般总希望一个好的预测，其均方预测

误差应达到最小，即

$$E[Y-M(X)]^2=\min_L E[Y-L(X)]^2 \tag{1}$$

其中 min 是对一切 x 的（可测）函数 $L(x)$ 取极小，对此有

定理 1 当 $M(X)$ 取作为条件数学期望

$$M(X)=E[Y|X] \tag{2}$$

时，使得(1)式成立．即

$$\sigma^2_{Y.X}\triangleq E[Y-E[Y|X]]^2=\min_L E[Y-L(X)]^2 \tag{3}$$

且 $M(X)$ 与 Y 具有最大相关，即

$$\rho(Y,\ M(X))=\max_L \rho(Y,\ L(X)) \tag{4}$$

［证明］（仅对连续型情形给出）

设 $(X,\ Y)$ 的分布密度是 $f(x,\ y)$，X 的边际分布密度是 $f_1(x)$，Y 关于 X 的条件分布密度是

$$f(y|x)=\begin{cases}\dfrac{f(x,\ y)}{f_1(x)}, & \text{当 } f_1(x)\neq 0\\ 0, & \text{当 } f_1(x)=0\end{cases}$$

则 Y 关于 X 的条件期望是

$$M(x)\triangleq E(Y|x)=\int yf(y|x)dy$$

$$E[Y-L(X)]^2=\iint\cdots\int[y-L(x)]^2f(x,\ y)dx\,dy$$

$$=\iint\cdots\int[y-M(x)]^2f(x,\ y)dx\,dy$$

$$+2\iint\cdots\int[y-M(x)][M(x)-L(x)]f(x,\ y)dx\,dy$$

$$+\iint\cdots\int[M(x)-L(x)]^2f(x,\ y)dx\,dy$$

由于

$$\iint\cdots\int[y-M(x)][M(x)-L(x)]f(x,\ y)dx\,dy$$

$$=\iint\cdots\int[y-M(x)][M(x)-L(x)]f(y|x)f_1(x)dy\,dx$$

$$=\int\cdots\int[M(x)-L(x)]f_1(x)\left[\int yf(y|x)dy-M(x)\right]dx$$

$$=0 \tag{5}$$

因而

$$E[Y-L(X)]^2=E[Y-M(X)]^2+E[M(X)-L(X)]^2 \tag{6}$$

(6) 式右边第一项与 $L(X)$ 无关，第二项大于等于零．它等于零的充要条件是

$$M(X)=L(X)\qquad a.s.$$

它表示当 $L(X)\overset{a.s.}{=}M(X)$ 时，$E[Y-L(X)]^2$ 达到最小值 $E[Y-M(X)]^2$.

在(6)式中取 $L(X)=EY$，并注意到 $E[M(X)]=EY$，有

$$D(Y)=D[Y-M(X)]+D(M(X))=\sigma^2_{Y.X}+D(M(X)) \tag{7}$$

考察

$$\operatorname{cov}(Y,\ L(X))=\operatorname{cov}(M(X),\ L(X))$$

$$+\iint\cdots\int\{[Y-E(Y)]-[M(X)-E(M(X))]\}$$

$$\times[L(X)-E(L(X))]f(x,\ y)dx\,dy$$

$$=\operatorname{cov}(M(X),\ L(X))$$

特别取 $L(X)=M(X)$ 得

$$\operatorname{cov}(Y,\ M(X))=\operatorname{cov}(M(X),\ M(X))$$

$$=D[M(X)]\geqslant 0$$

$$\rho^2(Y, L(X)) \triangleq \frac{\operatorname{cov}^2(Y, L(X))}{D(Y)D[L(X)]}$$

$$= \frac{\operatorname{cov}^2(M(X), L(X))}{D(Y)D[L(X)]}$$

$$= \frac{\operatorname{cov}^2(M(x), L(X))}{D[M(X)]D[L(X)]} \cdot \frac{D[M(X)]}{D(Y)} \cdot \frac{D[M(X)]}{D[M(X)]}$$

$$= \rho^2(M(X), L(X)) \cdot \rho^2(Y, M(X))$$

$$\leqslant \rho^2(Y, M(X))$$

所以 $$|\rho(Y, L(X))| \leqslant \rho(Y, M(X))$$

由 L 的任意性，因此(4)式成立.

上式"="成立的充要条件是

$$|\rho(M(X), L(X))| = 1$$

即 L 是 M 的线性函数.

在统计学上，我们称

$$y = M(x) = E[Y \mid X = x] \tag{8}$$

为 Y 关于 X 的**回归曲线**.

记

$$\eta_{Y,X} \triangleq \sqrt{\frac{D[M(X)]}{\sigma_Y^2}} = \rho(Y, M(X)) \tag{9}$$

其中 $\sigma_Y^2 \triangleq D(Y)$、$\eta_{Y,X}$ 称为**相关比**. 由(7)式得到

$$\eta_{Y,X} = \sqrt{1 - \frac{\sigma_{Y,X}^2}{\sigma_Y^2}} \tag{10}$$

由此可见，当 $\eta_{Y,X}$ 趋近于 1 时，预测误差方差 $\sigma_{Y,X}^2$ 趋近于 0，即预测误差趋近于 0，此时 $Y \approx M(X)$. 而当 $\eta_{Y,X}$ 接近于 0 时，$\sigma_{Y,X}$ 接近于 σ_Y^2，这说明运用 X 不能压缩 Y 的方差，所以相关比 $\eta_{Y,X}$ 提供了一个度量 Y 与 $X = (X_1, \cdots, X_m)^\tau$ 之间的联结程度的尺度，或者说是 Y 的预测精度的尺度.

[例 1] 设(X, Y)服从二元正态分布，分布密度是

$$f(x, y)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x_1-\mu_1)^2}{\sigma_1^2}\right.\right.$$
$$\left.\left.-\frac{2\rho(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2}+\frac{(y-\mu_2)^2}{\sigma_2^2}\right]\right\}$$

则容易计算得

$$f_1(x)=\frac{1}{\sqrt{2\pi}\,\sigma_1}\exp\left\{-\frac{(x-\mu_1)^2}{2\sigma_1^2}\right\}$$

$$f(y|x)=\frac{1}{\sigma_2\sqrt{2\pi}\sqrt{1-\rho^2}}\exp\left\{-\frac{1}{2\sigma_2^2(1-\rho^2)}\right.$$
$$\left.\times\left[y-\mu_2-\frac{\rho\sigma_2}{\sigma_1}(x-\mu_1)\right]^2\right\}$$

$$M(x)=\int yf(y|x)dy=\mu_2+\frac{\rho\sigma_2}{\sigma_1}(x-\mu_1)$$

所以 Y 关于 X 的回归曲线是一条直线,

$$y=\mu_2+\frac{\rho\sigma_2}{\sigma_1}(x-\mu_1) \tag{11}$$

二、线性回归

由例 1 可见, 对于正态随机变量, Y 关于 X 的回归曲线是直线, 即回归函数 $M(x)$ 是 x 的线性函数. 大家知道, 正态变量在实际问题中最常见. 另外, 对于非正态变量, 回归曲线一般不是线性函数, 但此时要从观察的数据资料来寻找回归函数十分困难, 因此也常常要求寻找一个线性函数

$$g(x)=g(x_1, \cdots, x_m)=\alpha+\beta_1x_1+\cdots+\beta_mx_m$$

使得 $g(X)$ 与 Y 的均方差异达到最小, 即

$$E\{|Y-(\alpha+\beta_1X_1+\cdots+\beta_mX_m)|^2\}=\min \tag{12}$$

其中 min 是对 X 的一切线性函数取极小.

满足(12)式的线性函数通常称为 Y 关于 X 的回归直线. 下面我们来求解回归直线的系数 α 和 $\beta^\tau=(\beta_1, \cdots, \beta_m)$.

记

$$L(\alpha,\ \beta)\triangleq E\{|Y-(\alpha+\beta_1X_1+\cdots+\beta_mX_m|^2\}$$
$$=b^2+\beta^\tau R_{XX}\beta-2\beta^\tau R_{XY}+D(Y) \tag{13}$$

其中 $b\triangleq EY-(\alpha+\beta_1EX_1+\cdots+\beta_mEX_m)$

$$R_{XX}\triangleq E\{(X-EX)(X-EX)^\tau\}$$
$$=\begin{pmatrix} D(X_1) & \text{cov}(X_1,\ X_2) & \cdots & \text{cov}(X_1, X_m) \\ \cdots & \cdots & \cdots & \cdots \\ \text{cov}(X_m,\ X_1) & \text{cov}(X_m, X_2) & \cdots & D(X_m) \end{pmatrix}_{m\times m}$$

$$R_{XY}\triangleq\begin{pmatrix} \text{cov}(Y,\ X_1) \\ \vdots \\ \text{cov}(Y,\ X_m) \end{pmatrix}_{m\times 1}$$

用微分方法求 $L(\alpha,\ \beta)$ 的极小值，得方程组

$$\begin{cases} b=0 \\ R_{XX}\beta=R_{XY} \end{cases} \tag{14}$$

解之得线性回归系数

$$\alpha^*=E(Y)-\beta_1^*E(X_1)-\cdots-\beta_m^*E(X_m)$$
$$=E(Y)-\beta^{*\tau}E(X) \tag{15}$$
$$\beta^*=R_{XX}^{-1}R_{XY}$$

这里假定 R_{XX}^{-1} 存在.

因此 Y 关于 X 的回归直线是

$$y=\alpha^*+\beta^{*\tau}x=\alpha^*+\beta_1^*x_1+\beta_2^*x_2+\cdots+\beta_m^*x_m \tag{16}$$

其中 α^*, β^* 由(15)式给定，而最小线性预测误差方差是

$$\sigma_{Y,X}^{*2}\triangleq\min_{\alpha,\beta} E\{[Y-(\alpha+\beta^\tau X)]^2\}$$
$$=E\{[Y-(\alpha^*+\beta^{*\tau}X)]^2\}$$
$$=\sigma_Y^2-R_{XY}^\tau R_{XX}^{-1}R_{XY} \tag{17}$$

由(15)、(16)、(17)式可见，对于线性回归，它仅仅是由 Y 和 X 的均值、方差、协方差(二阶矩)所决定，而并不需要知道分布的

其它知识.

类似于定理 1, 有

定理 2 对于 Y 关于 X 的线性回归函数 $\alpha^*+\beta^{*\tau}X$, 有

(1) 它是 Y 的均方误差最小的线性预测函数;

(2) 它也是与 Y 具有最大相关的线性函数.

[证明] (1) 是显然的, 现证(2).

对任一 X 的线性函数 $\alpha+\beta^\tau X$ 有

$$\mathrm{cov}(Y,\ \alpha+\beta^\tau X)=\mathrm{cov}(Y,\ \beta^\tau X)=\beta^\tau R_{XY}=\beta^\tau R_{XX}\beta^*$$

$$\mathrm{cov}(Y,\ \alpha^*+\beta^{*\tau}X)=\beta^{*\tau}R_{XX}\beta^*=D(X^\tau\beta^*)\geqslant 0$$

由柯西-许瓦兹不等式, 得

$$(\beta^\tau R_{XX}\beta^*)^2\leqslant(\beta^\tau R_{XX}\beta)\cdot(\beta^{*\tau}R_{XX}\beta^*)$$

$$\rho^2(Y,\ \alpha+\beta^\tau X)\triangleq\frac{\mathrm{cov}^2(Y,\ \alpha+\beta^\tau X)}{\sigma_Y^2\cdot D(\alpha+\beta^\tau X)}=\frac{(\beta^\tau R_{XX}\beta^*)^2}{\sigma_Y^2\cdot(\beta^\tau R_{XX}\beta)}$$

$$\leqslant\frac{(\beta^{*\tau}R_{XX}\beta^*)}{\sigma_Y^2}=\frac{\mathrm{cov}^2(Y,\ \alpha^*+X^\tau\beta^*)}{\sigma_Y^2 D(\alpha^*+X^\tau\beta^*)}$$

$$\triangleq\rho^2(Y,\ \alpha^*+X^\tau\beta^*)$$

所以 $$|\rho(Y,\ \alpha+\beta^\tau X)|\leqslant\rho(Y,\ \alpha^*+\beta^{*\tau}X)$$

由 α 和 β 的任意性, 得知线性回归函数是与 Y 具有最大相关的线性函数.

记 $$\rho_{Y,X}\triangleq\rho_{Y,(X_1,\cdots,X_m)}\triangleq\rho(Y,\ \alpha^*+X^\tau\beta^*)$$

$$=(R_{XY}^\tau R_{XX}^{-1}R_{XY})^{\frac{1}{2}}/\sigma_Y \tag{18}$$

$\rho_{Y,X}$ 称为复相关系数. 它指出了 Y 与 X 之间的线性相关程度. 如果 $\rho_{Y,X}=1$, 则说明 Y 是 $X_1,\ \cdots,\ X_m$ 的线性函数(在概率 1 意义下), 反之亦然.

显然

$$\eta_{Y,X}\geqslant\rho_{Y,X} \tag{19}$$

对于正态分布有

$$\eta_{Y,X}=\rho_{Y,X}.$$

在大量的实际问题中，Y 和 X 的均值、方差和协方差全部是未知的，也就是说在方程(15)或(16)式中的 EY，EX，R_{XX}，R_{XY} 都是未知的．此时 α^*，β^* 就无法求解． 在这种情况下，往往就要求我们通过$(X_1, \cdots, X_m, Y)$的观察值，来估计 α^* 和 β^*．

设有 n 个观察值，是 $\{x_{1k}, x_{2k}, \cdots, x_{mk}, y_k\}$，$k=1, 2, \cdots, n$．第一步我们采用矩方法估计 X 和 Y 的前二阶矩，记

$$\bar{y} \triangleq \frac{1}{n}\sum_{k=1}^{n} y_k$$

$$\bar{x}_i \triangleq \frac{1}{n}\sum_{k=1}^{n} x_{ik}, \quad i=1, \cdots, m$$

$$S_{yx_i}^2 \triangleq \frac{1}{n}\sum_{k=1}^{n} (y_k-\bar{y})(x_{ik}-\bar{x}_i), \quad i=1, \cdots, m$$

$$S_{x_ix_j}^2 \triangleq \frac{1}{n}\sum_{k=1}^{n} (x_{ik}-\bar{x}_i)(x_{jk}-\bar{x}_j), \quad i, j=1, \cdots, m$$

$$S_{yy}^2 \triangleq \frac{1}{n}\sum_{k=1}^{n} (y_k-\bar{y})^2$$

用 $\bar{y}$ 和 $\bar{x}_i$ 分别作为均值 EY 和 EX_i 的估计值，用

$$S_{XX} \triangleq \begin{pmatrix} S_{x_1x_1}^2 & S_{x_1x_2}^2 & \cdots & S_{x_1x_m}^2 \\ \cdots & \cdots & \cdots & \cdots \\ S_{x_mx_1}^2 & S_{x_mx_2}^2 & \cdots & S_{x_mx_m}^2 \end{pmatrix}$$

$$S_{YX} \triangleq \begin{pmatrix} S_{yx_1}^2 \\ \vdots \\ S_{yx_m}^2 \end{pmatrix}$$

分别作为 R_{XX}，R_{YX} 的估计值． 代入(15)式就得到线性回归系数 α^* 和 β^* 的估计 $\hat{\alpha}$ 和 $\hat{\beta}$ 为

$$\hat{\alpha}=\bar{y}-(\bar{x}_1, \cdots, \bar{x}_m)\hat{\beta}=\bar{y}-\hat{\beta}_1\bar{x}_1-\cdots-\hat{\beta}_m\bar{x}_m \qquad (20)$$

$$\hat{\beta}=S_{XX}^{-1}S_{YX}$$

又可写为：

$$\begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \end{pmatrix} = (X^{\tau}X)^{-1}X^{\tau}Y \tag{21}$$

其中

$$X \triangleq \begin{pmatrix} 1 & x_{11} & x_{21} & \cdots & x_{m1} \\ 1 & x_{12} & x_{22} & \cdots & x_{m2} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & x_{1n} & x_{2n} & \cdots & x_{mn} \end{pmatrix}_{n\times(m+1)}$$

$$Y \triangleq \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}_{n\times 1}$$

类似地，复相关系数的估计值是

$$\hat{\rho}^2_{YX} = (S^{\tau}_{YX}S^{-1}_{XX}S_{YX})/S^2_{yy}$$

$$= 1 - \frac{S^2_{yy} - S^{\tau}_{YX}S^{-1}_{XX}S_{YX}}{S^2_{yy}}$$

$$= 1 - \frac{Y^{\tau}Y - \hat{\beta}^{\tau}S_{YX}}{Y^{\tau}Y - n(\bar{y})^2} = 1 - \frac{Y^{\tau}Y - \begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \end{pmatrix}^{\tau}(X^{\tau}Y)}{Y^{\tau}Y - n(\bar{y})^2}$$

$$= 1 - \frac{R^2_0}{\sum\limits_{i=1}^{n}(y_i - \bar{y})^2} = 1 - \frac{R^2_0/n}{S^2_{yy}}$$

$$\hat{\rho}_{Y,X} = \sqrt{1 - \frac{R^2_0/n}{S^2_{yy}}} \tag{22}$$

其中 $R^2_0 = Y^{\tau}Y - \begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \end{pmatrix}^{\tau}(X^{\tau}Y)$，同(1.21)式．而 R^2_0/n 恰好是线性预测误差方差 $\sigma^{*2}_{Y,X}$（由(17)式给出）的矩估计．

比较公式(1.12)和(21)，两者在形式上是完全一致的．但是这两个公式的含义和推导过程却是不同的．

关于回归分析的含义说法不一，在有些教科书和文献中，把我们在§1中所讨论有关线性模型的统计分析称为回归分析，并类

似地定义偏相关系数(按 (22) 式计算) 等概念. 在本书中我们将 X 为随机变量时的有关统计分析称为回归分析. 虽然关于回归分析的定义很不一致, 但是由于求线性模型的未知参数的估计的公式(1.12) 和在回归分析中求线性回归系数估计的公式 (21) 在形式上是完全一致的. 因此从计算角度讲, 这一定义的混淆关系不大.

三、附加结合的度量方法

下面我们来研究当加入新的变量 $X_{m+1},\cdots,X_{m+p}$ 到原来的自变量 $X_1,\cdots,X_m$ 上去以后,对 Y 的预测精度的影响.

由(8)式和(10)式可见,通过变量 $X_1,\cdots,X_m$, 对 Y 的最佳预测的均方误差是

$$\sigma^2_{Y(X_1,\cdots,X_m)}=(1-\eta^2_{Y(X_1,\cdots,X_m)})\sigma^2_Y$$

而当加入新变量 $X_{m+1},\cdots,X_{m+p}$ 后,即运用变量 $X_1,\cdots,X_m,X_{m+1},\cdots,X_{m+p}$ 对 Y 进行预测,此时 Y 的最佳预测的均方误差是

$$\sigma^2_{Y(X_1,\cdots,X_{m+p})}=(1-\eta^2_{Y(X_1,\cdots,X_{m+p})})\sigma^2_Y$$

因此, 当加入新变量 $X_{m+1},\cdots,X_{m+p}$ 后,对 Y 的预测的均方误差的相对压缩比是

$$\eta^2_{Y(X_{m+1},\cdots,X_{m+p})\cdot(X_1,\cdots,X_m)}\triangleq\frac{\sigma^2_Y[1-\eta^2_{Y(X_1,\cdots,X_m)}]-\sigma^2_Y[1-\eta^2_{Y(X_1,\cdots,X_{m+p})}]}{\sigma^2_Y[1-\eta^2_{Y(X_1,\cdots,X_m)}]}$$

$$=\frac{\eta^2_{Y(X_1,\cdots,X_{m+p})}-\eta^2_{Y(X_1,\cdots,X_m)}}{1-\eta^2_{Y(X_1,\cdots,X_m)}}\tag{23}$$

$\eta_{Y(X_{m+1},\cdots,X_{m+p})\cdot(X_1,\cdots,X_m)}$ 称为偏相关比. 它与相关比不同,相关比是衡量 Y 与 X 之间直接结合程度大小的尺度, 而偏相关比是衡量消除了 $X_1,\cdots,X_m$ 的影响后, Y 与 $X_{m+1},\cdots,X_{m+p}$ 的结合程度,即附加结合程度大小的尺度.

如果我们仅研究线性预测,类似地运用(17)式和(18)式,此时对 Y 的最佳线性预测的平方平均误差的相对压缩比是

$$\rho^2_{Y(X_{m+1},\cdots,X_{m+p})\cdot(X_1,\cdots,X_m)} \triangleq \frac{\rho^2_{Y(X_1,\cdots,X_{m+p})} - \rho^2_{Y(X_1,\cdots,X_m)}}{1-\rho^2_{Y(X_1,\cdots,X_m)}} \tag{24}$$

$\rho_{Y(X_{m+1},\cdots,X_{m+p})\cdot(X_1,\cdots,X_m)}$ 称为 Y 与 X_{m+1}, $\cdots$, X_{m+p} 之间的偏复相关系数. 它表示消除了 X_1, $\cdots$, X_m 的影响之后, Y 与 X_{m+1}, $\cdots$, X_{m+p} 之间的线性结合程度.

在实用上, 对 $\rho_{Y(X_{m+1})\cdot(X_1,\cdots,X_m)}$ 特别感到兴趣. 因为它可以用来推断新加入 X_{m+1} 进去后是否有意义, 即 X_{m+1} 对提高 Y 的预测精度是否重要.

§3. 方差分析

在工农业生产及产品试制中, 常常需要分析哪几种因素对生产的质量和产量起显著影响, 并希望知道起显著作用的因素在什么时候, 对生产起最好的影响.

例如在农业科学试验中, 为了提高农作物的收获量, 因地制宜地选择品种. 常常需要比较不同品种的种子, 施不同种类、不同数量的肥料对农作物收获量的影响, 并从中找出最适宜于该地区的作物品种、肥料的种类和数量, 以提高单位面积的产量.

又如在橡胶配方中, 往往需要了解不同种类的促进剂, 不同份量的氧化锌和硫化条件对所得橡胶制品的质量的影响, 从中找出这些条件怎样搭配最为适宜, 使产品的质量尽可能地提高等等.

为了解决这类问题, 一般需要做下面两步工作. 第一是设计一个试验, 使得这个试验一方面能很好地反映出我们所感兴趣的因素的作用, 另一方面试验的次数要尽可能地少, 节约人力、物力和时间. 其次是如何充分地利用试验结果的信息, 对我们所关心的事物(因素的影响)作出合理的推断. 前者通常称为试验设计, 后者最常用的统计方法是本节所要介绍的方差分析①.

① 试验设计可参阅中国科学院数学研究所统计组,《常用数理统计方法》, 1973.

方差分析首先被应用于农业试验，它在农业、工业，生物学和医学等许多方面均具有广泛的应用.

一、一种方式分组的数据

先分析一个例子

[例 1] 某生产队为了比较四种不同的肥料对某农作物收获量的影响．进行了下面的试验．他们选用一块肥沃程度比较均匀的土地，并将这块土地按表 1 的方式等分成十六小块．同时为了尽可能减少土地原有肥沃程度差异的影响，采用表 1 的方式安排试验．

表 1

A_1	A_2	A_3	A_4
A_2	A_3	A_4	A_1
A_3	A_4	A_1	A_2
A_4	A_1	A_2	A_3

在表 1 中，A_i 表示在这一小块土地上施第 i 种肥料．显然施每种肥料的各有四小块土地．试验结果——作物收获量由表 2 给出.

表 2

肥料种类 (A_i)	收获量 (x_{ij})				平均收获量 ($\bar{x}_i$)
A_1	98	96	91	66	87.75
A_2	60	69	50	35	53.50
A_3	79	64	81	70	73.50
A_4	90	70	79	88	81.75

我们希望通过表 2 给出的数据(试验结果)来推断：肥料对该作物的收获量有否显著影响？如影响显著，那么施哪一种肥料为好.

在统计学上，我们通常称因素的不同状态为水平．在例 1 中，

考虑的是一个因素——肥料．这因素具有 4 个水平——A_1, A_2, A_3, A_4.

从例 1 中看到：

(1) 虽然所施的肥料相同，其它生产条件也一样，但相同面积土地上的收获量是不相等的，这说明该作物的收获量是一个随机变量.

(2) 由于在这里我们要研究四种不同肥料的影响，从表 2 右边所列出的平均收获量可以看出肥料对收获量是有影响的．因此不能把不同种类肥料所得的收获量视为从同一母体中所抽得的子样，而应视为是分别从四个母体 $\mathbb{X}_1$, $\mathbb{X}_2$, $\mathbb{X}_3$, $\mathbb{X}_4$ 中抽得的 4 个容量为 4 的子样.

现在判断肥料对作物产量影响的问题，就是要辨别收获量之间的差异主要是由抽样误差造成的还是由肥料的影响所造成的．这一问题可归结为判断四个母体是否具有相同分布的问题.

由于在实际中，我们经常遇到的是具有正态分布的母体．同时在安排试验时，除了我们所关心的因素(如在例 1 中的肥料)外，其它试验条件总是尽可能做到一致，这就使得我们可以认为每个母体的方差是相同的，因此在这里，推断几个母体是否具有相同分布的问题就简化为检验几个具有相同方差的正态母体，其均值是否相等的问题.

对于例 1，也就是假定

(1) $\mathbb{X}_1$, $\mathbb{X}_2$, $\mathbb{X}_3$, $\mathbb{X}_4$ 是相互独立的正态母体，分别服从于 $N(\mu_i, \sigma^2)$ 分布，$i=1, 2, 3, 4$.

(2) X_{ij} $j=1, 2, 3, 4$, $i=1, 2, 3, 4$ 是分别从母体 $\mathbb{X}_i$ 中所抽得的简单随机子样.

要求检验假设 H_0: $\mu_1=\mu_2=\mu_3=\mu_4$.

下面我们对更一般的问题推导统计方法.

设 p 个相互独立的正态母体 $\mathbb{X}_i$, $i=1, \cdots, p$. $\mathbb{X}_i \sim N(\mu_i, \sigma^2)$ 分布．记

$$X_{i1}, \cdots, X_{in_i}$$

是从第 i 个母体 $\mathbb{X}_i$ 中抽得的容量为 n_i 的简单随机子样，要求检验假设

$$H_0: \mu_1 = \mu_2 = \cdots = \mu_p \tag{1}$$

在上述的假定之下，如果记

$$n \triangleq \sum_{i=1}^{p} n_i$$

$$W \triangleq \begin{pmatrix} X_{11} \\ \vdots \\ X_{1n_1} \\ X_{21} \\ \vdots \\ X_{2n_2} \\ \vdots \\ X_{p1} \\ \vdots \\ X_{pn_p} \end{pmatrix}_{n\times 1}, \quad L \triangleq \begin{pmatrix} 1 & & & \\ \vdots & & & \\ 1 & & & \\ & 1 & & \\ & \vdots & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \\ & & & \vdots \\ & & & 1 \end{pmatrix}_{n\times p} \begin{matrix} \left.\vphantom{\begin{matrix}1\\1\\1\end{matrix}}\right\} n_1 \\ \left.\vphantom{\begin{matrix}1\\1\\1\end{matrix}}\right\} n_2 \\ \\ \left.\vphantom{\begin{matrix}1\\1\\1\end{matrix}}\right\} n_p \end{matrix}$$

$$\beta \triangleq \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{pmatrix}_{p\times 1}$$

则 W 为线性模型 $(W, L\beta, \sigma^2 I_n)$. 因此它可以应用关于线性模型的最小二乘法理论解决，但由于在方差分析中的模型，矩阵 L 经常是降秩的，因此 §1 的讨论不作进一步推广便不能直接应用，在这里我们将采用直观的离差分解的方法处理。

记

$$\overline{X}_i \triangleq \frac{1}{n_i}\sum_{j=1}^{n_i} X_{ij} \tag{2}$$

$$\overline{X} \triangleq \frac{1}{n}\sum_{i=1}^{p}\sum_{j=1}^{n_i} X_{ij} = \frac{1}{n}\sum_{i=1}^{p} n_i\overline{X}_i \tag{3}$$

$\overline{X}_i$ 是从第 i 个母体抽得的子样的平均值，它通常被称为组平均值，而 $\overline{X}$ 称为总平均值，n 是从 p 个母体抽得的子样的总容量.

由(2)，(3)可以推得

$$\sum_{i=1}^{p}\sum_{j=1}^{n_i}(X_{ij}-\overline{X}_i)(\overline{X}_i-\overline{X}) = 0$$

由此得到

$$\begin{aligned} Q &= \sum_{i=1}^{p}\sum_{j=1}^{n_i}(X_{ij}-\overline{X})^2 \\ &= \sum_{i=1}^{p}\sum_{j=1}^{n_i}[(X_{ij}-\overline{X}_i)+(\overline{X}_i-\overline{X})]^2 \\ &= \sum_{i=1}^{p}\sum_{j=1}^{n_i}(X_{ij}-\overline{X}_i)^2+\sum_{i=1}^{p} n_i(\overline{X}_i-\overline{X})^2 \\ &= Q_1+Q_2 \end{aligned} \tag{4}$$

其中

$$\begin{aligned} Q_1 &\triangleq \sum_{i=1}^{p}\sum_{j=1}^{n_i}(X_{ij}-\overline{X}_i)^2 \\ Q_2 &\triangleq \sum_{i=1}^{p} n_i(\overline{X}_i-\overline{X})^2 \end{aligned} \tag{5}$$

从 (4) 式看到，Q 是所有的观察资料 X_{ij} 与总平均数 $\overline{X}$ 的差异的平方和，是描写所得全部数据离散程度的一个指标. 它可以分解为两项，第一项 Q_1 是每个观察数据与其组平均值的差异的平方和，它反映了观察 X_{ij} 抽样误差的大小程度，而第二项 Q_2 是组平均与总平均的差异的平方和，它在一定程度上反映了各母体均值 μ_i 之间的差异程度. 为更清楚地说明这个事实，分别计算它们的数学期望，得

$$E(Q_1)=\sum_{i=1}^{p}E\left\{\sum_{j=1}^{n_i}(X_{ij}-\overline{X}_i)^2\right\}$$
$$=\sum_{i=1}^{p}(n_i-1)\sigma^2=(n-p)\sigma^2$$
$$E(Q_2)=\sum_{i=1}^{p}n_iE(\overline{X}_i-\overline{X})^2$$
$$=(p-1)\sigma^2+\sum_{i=1}^{p}n_i(\mu_i-\mu)^2$$

其中
$$\mu\triangleq\frac{1}{n}\sum_{i=1}^{p}n_i\mu_i$$

记
$$S_1^2\triangleq\frac{Q_1}{n-p}$$
$$S_2^2\triangleq\frac{Q_2}{p-1} \tag{6}$$

则
$$E(S_1^2)=\sigma^2$$
$$E(S_2^2)=\sigma^2+\frac{1}{p-1}\sum_{i=1}^{p}n_i(\mu_i-\mu)^2 \tag{7}$$

由此可见，不管对于 μ_i 的假设如何，S_1^2 是 σ^2 的一个无偏估计，而 S_2^2 仅当假设 $H_0(\mu_1=\mu_2=\cdots=\mu_p)$ 成立时，才是 σ^2 的一个无偏估计，否则它的期望值要大于 σ^2. 这说明，比值

$$F=\frac{S_2^2}{S_1^2}=\frac{(n-p)Q_2}{(p-1)Q_1} \tag{8}$$

在假设 H_0 不成立时，有偏大的倾向.

现在，我们运用第二章的结果来讨论统计量 F 的分布。

当 H_0 成立时，$\mu_1=\mu_2=\cdots=\mu_p=\mu$ 由(4)式得到

$$\sum_{i=1}^{p}\sum_{j=1}^{n_i}(X_{ij}-\mu)^2=Q_1+Q_2+n(\overline{X}-\mu)^2$$

对于 Q_1，它有 p 个线性关系 $\sum_{j=1}^{n_i}(X_{ij}-\overline{X}_i)=0$，$i=1, \cdots, p$. 所以 Q_1 的秩是$(n-p)$.

对于 Q_2，它包含一个线性关系 $\sum_{i=1}^{p} n_i(\overline{X}_i-\overline{X})=0$，所以 Q_2 的秩是 $(p-1)$.

$n(\overline{X}-\mu)^2$ 的秩为 1.

这三项都是非负定二次型，而 $(n-p)+(p-1)+1=n$. 所以由定理 2.2.4 和定理 2.3.5，在假设 H_0 成立时，$\frac{Q_1}{\sigma^2}$ 和 $\frac{Q_2}{\sigma^2}$ 分别为相互独立的 $\chi^2(n-p)$ 和 $\chi^2(p-1)$ 分布. 因此由(8)式所给定的统计量 F 服从于自由度为 $(p-1, n-p)$ 的 F-分布. 于是，对于显著水平 α，可查表（见附表 4）求得 F_α，使 $P(F\geqslant F_\alpha)=\alpha$. 当 F 的观察值大于等于 F_α 时，拒绝假设 $H_0(\mu_1=\mu_2=\cdots=\mu_p)$，否则认为试验结果与假设 H_0 无显著差异.

为应用方便起见，常用下面的方差分析表

表 3 一种方式分组的方差分析表

方差来源	平方和	自由度	均方	F 值
因素的影响	$Q_2=\sum_{i=1}^{p} n_i(\overline{X}_i-\overline{X})^2$	$p-1$	$S_2^2=\frac{1}{p-1}Q_2$	$\frac{S_2^2}{S_1^2}$
误差	$Q_1=\sum_{i=1}^{p}\sum_{j=1}^{n_i}(X_{ij}-\overline{X}_i)^2$	$n-p$	$S_1^2=\frac{1}{n-p}Q_1$	
总和	$Q=\sum_{i=1}^{p}\sum_{j=1}^{n_i}(X_{ij}-\overline{X})^2$	$n-1$	$S^2=\frac{1}{n-1}Q$	

[例 1] （续）由表 2 可算得

$$n=16,\ p=4,\ n_i=4,\ i=1,\ 2,\ 3,\ 4$$

$$\bar{x}=74.125,\ s_2^2=892.55,\ s_1^2=141.625$$

$$F=\frac{892.55}{141.625}\approx 6.30$$

或用方差分析表，得

表 4

方差来源	平方和	自由度	均方	F 值
A的影响	2677.65	3	892.55	6.30
误差	1699.5	12	141.625	
总和	4377.15	15		

这里 F 的自由度是(3, 12)，对 $\alpha=0.01$，查表(附表4)得 $F_\alpha=5.95$. 现在 $F=6.30>5.95=F_\alpha$，所以拒绝假设 H_0，即认为肥料对该农作物的收获量有显著影响.

二、两种方式分组的数据

上面介绍了一个因素影响时的方差分析方法. 但在许多实际问题中，常常要同时研究几种因素的影响. 类似于例1，在农业试验中有时要同时研究几种不同品种的种子和几种不同种类的肥料对农作物收获量的影响. 这里就有两个因素，一个因素是种子的品种；另一个因素是肥料的种类. 它们两者同时影响着作物的收获量. 与前面一个因素影响的情况类似，我们希望通过试验选取使收获量达到最高的种子品种和肥料种类. 这里由于存在两个因素的影响，就产生了一个新问题：不同品种的种子和不同种类的肥料对收获量的联合影响是否正好是它们每个因素分别对收获量的影响的迭加呢？也就是说，是否会产生这样的情况，分别使收获量达到最高的种子品种与肥料类型搭配在一起会使收获量大幅度提高，或者反而使收获量下降，而看来不是最好(即收获量不是最高)的种子品种和肥料类型搭配在一起. 由于搭配得当而得到最高的收获量. 这种各个因素的不同水平的搭配所产生的新的影响在统计学上称为**交互作用**. 各因素之间是否存在交互作用这是多因素方差分析中产生的新问题. 这个问题在上面所讲的单因子方差分析中不存在. 由于多因素问题很复杂，而解题的思想和基本方法又类同，所以我们仅介绍二因子方差分析问题.

在两因子的方差分析中，只有当在每个因素的不同水平上进行重复观察时，才能分析出是否存在交互影响。为了更好地了解二因子的方差分析方法，我们分两步，先介绍没有重复观察时的方差分析，再介绍具有相等重复观察次数的方差分析方法。

设因素 A 有 p 个不同的水平

$$A_1, A_2, \cdots, A_p$$

因素 B 有 q 个不同的水平

$$B_1, B_2, \cdots, B_q$$

对每种情况 $A_i \times B_j$ 进行一次独立试验，共得 pq 个试验结果 X_{ij}，由下表给出

表 5

		B 因子				平均值 $\bar{X}_{k\cdot}$
		B_1	B_2	$\cdots$	B_q	
A 因子	A_1	X_{11}	X_{12}	$\cdots$	X_{1q}	$\bar{X}_{1\cdot}$
	A_2	X_{21}	X_{22}	$\cdots$	X_{2q}	$\bar{X}_{2\cdot}$
	$\vdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
	A_p	X_{p1}	X_{p2}	$\cdots$	X_{pq}	$\bar{X}_{p\cdot}$
平均值 $\bar{X}_{\cdot l}$		$\bar{X}_{\cdot 1}$	$\bar{X}_{\cdot 2}$	$\cdots$	$\bar{X}_{\cdot q}$	$\bar{X}$

其中

$$\bar{X}_{k\cdot} \triangleq \frac{1}{q}\sum_{l=1}^{q} X_{kl},\ k=1, \cdots, p$$

$$\bar{X}_{\cdot l} \triangleq \frac{1}{p}\sum_{k=1}^{p} X_{kl},\ l=1, \cdots, q \tag{9}$$

$$\bar{X} \triangleq \frac{1}{pq}\sum_{i=1}^{p}\sum_{j=1}^{q} X_{ij}$$

在这里，我们类似地假设 X_{kl} 是相互独立的 $N(\mu_{kl}, \sigma^2)$ 随机变量。也就是说 X_{kl} 是从分布为 $N(\mu_{kl}, \sigma^2)$ 的母体中抽得的子样，且是相互独立的。由于这里我们认为 A, B 两因素之间不存在交互作用，故假定其均值

$$\mu_{kl}=\mu+\alpha_k+\beta_l,\ k=1,\ \cdots p,\ l=1,\ \cdots,\ q \tag{10}$$

其中 α_k, β_l 满足

$$\sum_{k=1}^{p}\alpha_k=0,\quad \sum_{l=1}^{q}\beta_l=0 \tag{11}$$

这里参数 μ 是 pq 个母体的均值的平均，参数 α_k 表示因素 A 的各个不同水平的影响的大小，参数 β_l 表示因素 B 的各个不同水平影响的大小. 因此我们要判断因素 A 的影响是否显著就等价于要检验假设

$$H_{01}: \alpha_1=\alpha_2=\cdots=\alpha_p=0 \tag{12}$$

类似地，要判断因素 B 的影响是否显著就等价于检验假设

$$H_{02}: \beta_1=\beta_2=\cdots=\beta_q=0 \tag{13}$$

为了检验这些假设，我们将离差的总平方和 Q 进行分解:

$$\begin{aligned}Q&\triangleq\sum_{k=1}^{p}\sum_{l=1}^{q}(X_{kl}-\overline{X})^2=\sum_{k=1}^{p}\sum_{l=1}^{q}[(X_{kl}-\overline{X}_{k\cdot}-\overline{X}_{\cdot l}+\overline{X})\\&\quad+(\overline{X}_{\cdot l}-\overline{X})+(\overline{X}_{k\cdot}-\overline{X})]^2\\&=q\cdot\sum_{k=1}^{p}(\overline{X}_{k\cdot}-\overline{X})^2+p\sum_{l=1}^{q}(\overline{X}_{\cdot l}-\overline{X})^2\\&\quad+\sum_{k=1}^{p}\sum_{l=1}^{q}(X_{kl}-\overline{X}_{k\cdot}-\overline{X}_{\cdot l}+\overline{X})^2\\&=Q_1+Q_2+Q_3\end{aligned} \tag{14}$$

其中

$$\begin{aligned}Q_1&\triangleq q\sum_{k=1}^{p}(\overline{X}_{k\cdot}-\overline{X})^2\\Q_2&\triangleq p\sum_{l=1}^{q}(\overline{X}_{\cdot l}-\overline{X})^2\\Q_3&\triangleq\sum_{k=1}^{k}\sum_{l=1}^{q}(X_{kl}-\overline{X}_{k\cdot}-\overline{X}_{\cdot l}+\overline{X})^2\end{aligned} \tag{15}$$

不难计算，可以得到

$$EQ_1=(p-1)\sigma^2+q\sum_{k=1}^{p}\alpha_k^2$$

$$EQ_2=(q-1)\sigma^2+p\sum_{l=1}^{q}\beta_l^2 \tag{16}$$

$$EQ_3=(p-1)(q-1)\sigma^2$$

令

$$S_1^2\triangleq\frac{1}{p-1}Q_1$$

$$S_2^2\triangleq\frac{1}{q-1}Q_2 \tag{17}$$

$$S_3^2\triangleq\frac{1}{(p-1)(q-1)}Q_3$$

则

$$ES_1^2=\sigma^2+\frac{q}{p-1}\sum_{k=1}^{p}\alpha_k^2$$

$$ES_2^2=\sigma^2+\frac{p}{q-1}\sum_{l=1}^{q}\beta_l^2 \tag{18}$$

$$ES_3^2=\sigma^2$$

因此，与一个因子的方差分析方法一样，可采用统计量

$$F_A=\frac{S_1^2}{S_3^2}=\frac{Q_1}{Q_3}\cdot\frac{(p-1)(q-1)}{p-1} \tag{19}$$

和

$$F_B=\frac{S_2^2}{S_3^2}=\frac{Q_2}{Q_3}\cdot\frac{(p-1)(q-1)}{q-1} \tag{20}$$

统计量 F_A 当 H_{01} 不成立时，有偏大的趋势，因此可以运用 F_A 对假设 H_{01} 进行检验；而统计量 F_B 当 H_{02} 不成立时，有偏大的趋势，因此可以运用 F_B 对假设 H_{02} 进行检验.

现在我们运用第二章的结果分析统计量 F_A 和 F_B 的分布. 在假设 H_{01} 和 H_{02} 成立时，$\mu_{kl}=\mu$，由(14)式得到

$$\sum_{k=1}^{p}\sum_{l=1}^{q}(X_{kl}-\mu)^2$$

$$=Q_1+Q_2+Q_3+pq(\overline{X}-\mu)^2$$

显然 Q_1, Q_2, Q_3 和 $pq(\overline{X}-\mu)^2$ 都是非负定二次型.

对于 Q_1，包含有一个线性关系 $\sum_{k=1}^{p}(\overline{X}_{k\cdot}-\overline{X})=0$，所以 Q_1 的秩是 $(p-1)$.

对于 Q_2，包含有一个线性关系 $\sum_{l=1}^{q}(\overline{X}_{\cdot l}-\overline{X})=0$，所以 Q_2 的秩是 $(q-1)$.

对于 Q_3，包含有线性关系 $\sum_{k=1}^{p}(X_{kl}-\overline{X}_{k\cdot}-\overline{X}_{\cdot l}+\overline{X})=0$，$l=1,\cdots,q$ 和 $\sum_{l=1}^{q}(X_{kl}-\overline{X}_{k\cdot}-\overline{X}_{\cdot l}+\overline{X})=0$，$k=1,\cdots,p$. 但是这 $(p+q)$ 个线性关系不是独立的. 因为 $\sum_{k=1}^{p}\sum_{l=1}^{q}(X_{kl}-\overline{X}_{k\cdot}-\overline{X}_{\cdot l}+\overline{X})=0$，所以 Q_3 的秩是 $pq-p-q+1=(p-1)(q-1)$.

$pq(\overline{X}-\mu)^2$ 的秩是 1.

而 $$(p-1)+(q-1)+(p-1)(q-1)+1=pq$$

运用定理2.2.4知道当假设 H_{01} 和 H_{02} 同时成立时，Q_1/σ^2，Q_2/σ^2 和 Q_3/σ^2 是相互独立的 χ^2-变量. 其自由度分别为 $(p-1)$，$(q-1)$ 和 $(p-1)(q-1)$.

实际上，可以证明当 H_{01} 成立时，Q_1/σ^2 和 Q_3/σ^2 为相互独立的 χ^2-变量. 当 H_{02} 成立时，Q_2/σ^2 和 Q_3/σ^2 为相互独立的 χ^2-变量. 因此，当 H_{01} 成立时由(19)式定义的统计量 F_A 服从 $F(p-1,(p-1)(q-1))$ 分布，可用它检验因素 A 的影响是否显著，即检验假设 H_{01}. 当 H_{02} 成立时，由(20)式定义的统计量 F_B 服从 $F(q-1,(p-1)(q-1))$ 分布，可用它检验因素 B 的影响是否显著，即检验假设 H_{02}. 为方便起见，常用下面的方差分析表.

表 6

方差来源	平 方 和	自由度	均方	F 值
A的影响	$Q_1=q\sum_{k=1}^{p}(\overline{X}_{k\cdot}-\overline{X})^2$	$p-1$	S_1^2	$F_A=S_1^2/S_3^2$
B的影响	$Q_2=p\sum_{l=1}^{q}(\overline{X}_{\cdot l}-\overline{X})^2$	$q-1$	S_2^2	$F_B=S_2^2/S_3^2$
误 差	$Q_3=\sum_{k=1}^{p}\sum_{l=1}^{q}(X_{kl}-\overline{X}_{k\cdot}-\overline{X}_{\cdot l}+\overline{X})^2$	$(p-1)(q-1)$	S_3^2	
总 和	$Q=\sum_{k=1}^{p}\sum_{l=1}^{q}(X_{kl}-\overline{X})^2$	$pq-1$		

[例 2] 为了研究酵解作用对血糖浓度的影响，从 8 名健康人中抽取了血液并制备成血滤液．每一个受试者的血滤液又分成 4 分．然后随机地把 4 分血滤液分别放置 0, 45, 90, 135 分钟，测定其中血糖浓度，测得数据如下：

表 7

		放置时间 (B)				$\bar{X}_{i\cdot}$
		0′	45′	90′	135′	
受试者编号 (A)	1	95	95	89	83	90.5
	2	95	94	88	84	90.25
	3	106	105	87	90	99.5
	4	98	97	95	90	95.0
	5	102	98	97	88	96.25
	6	112	112	101	94	104.75
	7	105	103	97	88	98.25
	8	95	92	90	80	89.25
$\bar{X}_{\cdot j}$		101	99.5	94.2	87.1	

试问：(1) 放置不同时间的血糖浓度的差别是否显著；(2) 不同受试者的血糖浓度的差别是否显著．

通过表 7 可算得对应的方差分析表．

表 8

方差来源	平方和	自由度	均方	F 值
受试者差异	806.3	7	115.2	28.8
放置时间影响	943.6	3	314.5	78.6
误差	84.1	21	4.0	
总和	1834.0	31		

取 $\alpha=0.01$．查 F-分布表(附表 4)．

对自由度(7, 21)得 $F_{0.01}=3.67$,

对自由度(3, 21)得 $F_{0.01}=4.90$. 显然

$$F_A=28.8>3.67=F_{0.01}$$

$$F_B=78.6\gg 4.90=F_{0.01}$$

所以可认为不同放置时间造成血糖浓度的差别是很显著的.不同受试者之间血糖浓度的差别也是显著的.

三、具有重复观察的两种方式分组的数据

在上面的讨论中，由于对 A, B 两因素的各种水平的组合仅进行一次观察，所以不能了解 A, B 两因素之间是否存在交互作用的影响. 而交互作用的影响正是单因素分析与多因素分析的本质区别所在. 为了考察交互作用的影响，对两个因素的各种水平 (A_k, B_l) 重复进行 c 次观察. 其观察值记为

$$X_{klj},\ j=1,\cdots,c,\ l=1,\cdots,q,\ k=1,\cdots,p$$

在这里我们假定:

(1) X_{klj}, $k=1,\cdots,p$, $l=1,\cdots,q$, $j=1,\cdots,c$, 相互独立，分别服从 $N(\mu_{kl},\sigma^2)$ 分布;

(2) $\mu_{kl}=\mu+\alpha_k+\beta_l+\delta_{kl}$

且
$$\sum_{k=1}^{p}\alpha_k=0,\quad \sum_{l=1}^{q}\beta_l=0 \tag{21}$$

$$\sum_{k=1}^{p}\delta_{kl}=\sum_{l=1}^{q}\delta_{kl}=0,\ k=1,\cdots,p,\ l=1,\cdots,q$$

其中 α_k, $k=1,\cdots,p$ 分别表示因素 A 的各个水平的影响. β_l, $l=1,\cdots,q$ 分别表示因素 B 的各个水平的影响. δ_{kl}, $k=1,\cdots,p$; $l=1,\cdots,q$ 分别表示因素 A、B 各个水平之间的交互作用的影响. 因此要判断因素 A、B 的影响及交互作用的影响是否显著分别等价于检验假设

$$H_{01}: \alpha_1=\alpha_2=\cdots=\alpha_p=0$$

$$H_{02}: \beta_1=\beta_2=\cdots=\beta_q=0 \tag{22}$$

$$H_{03}: \delta_{kl}=0,\ 对一切\ k=1,\cdots,p,\ l=1,\cdots,q$$

为了检验这些假设，类似地我们将离差的总平方和 Q 进行分解：

$$
\begin{aligned}
Q &= \sum_{k=1}^{p}\sum_{l=1}^{q}\sum_{j=1}^{c}(X_{klj}-\overline{X})^2 \\
&= \sum_{k=1}^{p}\sum_{l=1}^{q}\sum_{j=1}^{c}[(\overline{X}_{k\cdot\cdot}-\overline{X})+(\overline{X}_{\cdot l\cdot}-\overline{X}) \\
&\quad +(\overline{X}_{kl\cdot}-\overline{X}_{k\cdot\cdot}-\overline{X}_{\cdot l\cdot}+\overline{X})+(X_{klj}-\overline{X}_{kl\cdot})]^2 \\
&= Q_1+Q_2+Q_3+Q_4
\end{aligned}
\tag{23}
$$

其中

$$
\begin{aligned}
Q_1 &\triangleq qc\sum_{k=1}^{p}(\overline{X}_{k\cdot\cdot}-\overline{X})^2 \\
Q_2 &\triangleq pc\sum_{l=1}^{q}(\overline{X}_{\cdot l\cdot}-\overline{X})^2 \\
Q_3 &\triangleq c\sum_{k=1}^{p}\sum_{l=1}^{q}(\overline{X}_{kl\cdot}-\overline{X}_{k\cdot\cdot}-\overline{X}_{\cdot l\cdot}+\overline{X})^2 \\
Q_4 &\triangleq \sum_{k=1}^{p}\sum_{l=1}^{q}\sum_{j=1}^{c}(X_{klj}-\overline{X}_{kl\cdot})^2
\end{aligned}
\tag{24}
$$

$$\overline{X}_{kl\cdot}\triangleq\frac{1}{c}\sum_{j=1}^{c}X_{klj},\ k=1,\ \cdots,\ p,\ l=1,\ \cdots,\ q$$

$$\overline{X}_{k\cdot\cdot}\triangleq\frac{1}{q}\sum_{l=1}^{q}\overline{X}_{kl\cdot},\ k=1,\ \cdots,\ p$$

$$\overline{X}_{\cdot l\cdot}\triangleq\frac{1}{p}\sum_{k=1}^{p}\overline{X}_{kl\cdot},\ l=1,\ \cdots,\ q$$

$$\overline{X}\triangleq\frac{1}{pqc}\sum_{k=1}^{p}\sum_{l=1}^{q}\sum_{j=1}^{c}X_{klj}$$

类似地可以算得

$$
\left.
\begin{aligned}
E(Q_1) &= (p-1)\sigma^2+qc\sum_{k=1}^{p}\alpha_k^2 \\
E(Q_2) &= (q-1)\sigma^2+pc\sum_{l=1}^{q}\beta_l^2 \\
E(Q_3) &= (p-1)(q-1)\sigma^2+c\sum_{k=1}^{p}\sum_{l=1}^{q}\delta_{kl}^2 \\
E(Q_4) &= pq(c-1)\sigma^2
\end{aligned}
\right\}
\tag{25}
$$

令

$$\left.\begin{aligned} S_1^2 &\triangleq \frac{1}{p-1}Q_1, & S_2^2 &\triangleq \frac{1}{q-1}Q_2 \\ S_3^2 &\triangleq \frac{1}{(p-1)(q-1)}Q_3, & S_4^2 &\triangleq \frac{1}{pq(c-1)}Q_4 \end{aligned}\right\} \tag{26}$$

则有

$$\left.\begin{aligned} E(S_1^2) &= \sigma^2 + \frac{qc}{(p-1)}\sum_{k=1}^{p}\alpha_k^2 \\ E(S_2^2) &= \sigma^2 + \frac{pc}{(q-1)}\sum_{l=1}^{q}\beta_l^2 \\ E(S_3^2) &= \sigma^2 + \frac{c}{(p-1)(q-1)}\sum_{k=1}^{p}\sum_{l=1}^{q}\delta_{kl}^2 \\ E(S_4^2) &= \sigma^2 \end{aligned}\right\} \tag{27}$$

造统计量

$$F_A \triangleq S_1^2/S_4^2 \tag{28}$$

$$F_B \triangleq S_2^2/S_4^2 \tag{29}$$

$$F_{AB} \triangleq S_3^2/S_4^2 \tag{30}$$

显然由(28)式定义的统计量 F_A，当假设 H_{01} 不成立时，有偏大的趋势．故可以运用统计量 F_A 检验假设 H_{01}．

由(29)式定义的统计量 F_B，当假设 H_{02} 不成立时，有偏大的趋势，故可以运用统计量 F_B 检验假设 H_{02}．

由(30)式定义的统计量 F_{AB}，当假设 H_{03} 不成立时，有偏大的趋势，故可用来检验假设 H_{03}．

类似地运用定理 2.2.4 及定理 2.3.5 可以证明

Q_4/σ^2 服从 $\chi^2(pq(c-1))$ 分布．

当假设 H_{01} 成立时，Q_1/σ^2 服从 $\chi^2(p-1)$ 分布，且与 Q_4 独立，所以 F_A 服从 $F(p-1, pq(c-1))$ 分布．

当假设 H_{02} 成立时，Q_2/σ^2 服从 $\chi^2(q-1)$ 分布，且与 Q_4 独立，所以 F_B 服从 $F(q-1, pq(c-1))$ 分布．

当假设 H_{03} 成立时，Q_3/σ^2 服从 $\chi^2((p-1)(q-1))$ 分布，且

与 Q_4 独立，所以 F_{AB} 服从 $F((p-1)(q-1), pq(c-1))$ 分布。

为方便起见，常用下面的方差分析表。

表 9

方差来源	平方和	自由度	均方	F 值
A 的影响	$Q_1=qc\sum_{k=1}^{p}(\bar{X}_{k\cdot\cdot}-\bar{X})^2$	$p-1$	$S_1^2=\frac{Q_1}{p-1}$	$\frac{S_1^2}{S_4^2}$
B 的影响	$Q_2=pc\sum_{l=1}^{q}(\bar{X}_{\cdot l\cdot}-\bar{X})^2$	$q-1$	$S_2^2=\frac{Q_2}{q-1}$	$\frac{S_2^2}{S_4^2}$
交互影响 (A×B)	$Q_3=c\sum_{k=1}^{p}\sum_{l=1}^{q}(\bar{X}_{kl\cdot}-\bar{X}_{k\cdot\cdot}-\bar{X}_{\cdot l\cdot}+\bar{X})^2$	$(p-1)(q-1)$	$S_3^2=\frac{Q_3}{(p-1)(q-1)}$	$\frac{S_3^2}{S_4^2}$
误差	$Q_4=\sum_{k=1}^{p}\sum_{l=1}^{q}\sum_{j=1}^{c}(X_{klj}-\bar{X}_{kl\cdot})^2$	$pq(c-1)$	$S_4^2=\frac{Q_4}{pq(c-1)}$	
总和	$\sum_{k=1}^{p}\sum_{l=1}^{q}\sum_{j=1}^{c}(X_{klj}-\bar{X})^2$	$pqc-1$		

［例3］ 在某橡胶的配方中，考虑三种不同的促进剂，四种不同份量的氧化锌，同样的配方重复一次，测得 300% 的定伸强力如下：

表 10

促进剂 (A)	氧化锌 (B)			
	B_1	B_2	B_3	B_4
	定伸强力			
A_1	31, 33	34, 36	35, 36	39, 38
A_2	33, 34	36, 37	37, 39	38, 41
A_3	35, 37	37, 38	39, 40	42, 44

试问氧化锌、促进剂以及它们的交互作用对定伸强力有无显著影响？

由表 10 可算得相应的方差分析表。

表 11

方差来源	平方和	自由度	均　方	F 值
A	56.6	2	28.3	19.4
B	132.2	3	44.1	30.2
$A\times B$	4.7	6	0.8	0.55
误　差	17.5	12	1.46	
总　和	211.0	23		

取 $\alpha=0.01$，对自由度(2, 12)得 $F_\alpha=6.9$；对自由度 (3, 12) 得 $F_\alpha=6.0$；取 $\alpha=0.05$，对自由度 (6, 12) 得 $F_\alpha=3.0$（见附表4）.

表 11 中的 F 值与上述的临界值 F_α 进行比较，看出促进剂种类的影响和氧化锌总量的影响都是很显著的，而它们之间的交互作用认为可以忽略.

习　题

1．在考察硝酸钠 $NaNO_3$ 的可溶性程度时，对一列不同的温度(°C)观察它在 100 ml 的水中溶解的 $NaNO_3$ 的重量(g)，得观察结果如下：

温　度　(x_i)	0	4	10	15	21	29	36	51	68
$NaNO_3$ 的重量 (Y_i)	66.7	71.0	76.3	80.6	85.7	92.9	99.4	113.6	125.1

从经验与理论知道，Y 服从线性模型

$$Y=a+bx+e$$

试用最小二乘法估计未知参数 a，b 和 e 的方差 σ^2.

2．设

$$Y_1=a+e_1$$
$$Y_2=2a-b+e_2$$
$$Y_3=a+2b+e_3$$

其中 $e_i(i=1, 2, 3)$相互独立，且 $Ee_i=0$，$D(e_i)=\sigma^2$. 试求 a 和 b 的最小二

乘估计。

3. 设 $Y=\begin{pmatrix}Y_1\\Y_2\\Y_3\end{pmatrix}$ 服从线性模型 $(Y, X\beta, \sigma^2 I_3)$，

$$EY_i=\beta_0+\beta_1 x_i+\beta_2(3x_i^2-2),\ i=1, 2, 3$$

其中 $x_1=-1$, $x_2=0$, $x_3=1$. 试写出矩阵 X, 并求 β_0, β_1, β_2 的最小二乘估计.

4. 设 $Y=(Y_1, \cdots, Y_n)^\tau$ 服从线性模型 $(Y, X\beta, \sigma^2 I_n)$

$$EY_i=\beta_0+\beta_1 x_i,\ i=1, \cdots, n$$

试求 β_0 和 β_1 的最小二乘估计和 σ^2 的估计，并证明 β_0 和 β_1 不相关的充要条件是 $\bar{x}=0$.

5. 设 $Y=(Y_1, \cdots, Y_n)^\tau$ 服从线性模型 $(Y, X\beta, \sigma^2 I_n)$，其中 X 是 $n\times m$ $(n>m)$ 满列秩阵，β 是 $m\times 1$ 向量，Y 服从正态分布. 试求未知参数 β 和 σ^2 的最大似然估计，并与其最小二乘估计进行比较.

6. 某职工医院用光电比色计检验尿汞时，得尿汞含量 (mg/l) 与消光系数读数的结果如下：

尿汞含量 (x_i)	2	4	6	8	10
消光系数 (Y_i)	64	138	205	285	360

已知它们之间服从线性模型

$$EY_i=\beta_0+\beta_1 x_i$$

试求 β_0 和 β_1 的最小二乘估计，并检验 β_1 是否为零 $(\alpha=0.05)$？

7. 为了研究某种延迟血液凝固剂的效用，对注射了不同剂量该凝固剂的病员观察其血液凝固所需的时间 Y，得观察结果如下(这里采用 3 种不同的剂量，对每种剂量取 4 个观察)：

剂量 (x_i)	凝固时间 (Y_i)			
1.28	2.50	2.92	2.50	2.66
1.60	4.92	4.92	4.50	4.66
2.00	9.50	9.75	9.00	10.00

由于人们的个体差异、凝血时间 Y 是一个随机变量. 由经验知道，Y_i 服从 $N(\beta_0+\beta_1 x_i, \sigma^2)$ 分布. 试由此观察结果求 β_0, β_1 的最小二乘估计并检验 $\beta_1=0$ 是否为真 $(\alpha=0.05)$？ 如果 β_1 不为零，则求出它的 95% 的置信区间.

8. 设 Y 服从线性模型$(Y, X\beta, \sigma^2 I_n)$，$X$ 是 $n\times m$ $(n>m)$满列秩阵，如果分块表示 X 和 β 为

$$X\beta=(X_1, X_2)\begin{pmatrix}\beta_1\\ \beta_2\end{pmatrix}$$

证明 β_2 的最小二乘估计是

$$\hat{\beta}_2=[X_2^\tau X_2-X_2^\tau X_1(X_1^\tau X_1)^{-1}X_1^\tau X_2]^{-1}[X_2^\tau Y-X_2^\tau X_1(X_1^\tau X_1)^{-1}X_1^\tau Y]$$

并求 $\hat{\beta}_2$ 的协方差阵.

9. 设 Y 服从线性模型$(Y, X\beta, \sigma^2 I_n)$，且设

$$EY_i=\beta_0+\beta_1 x_i, \quad i=1, \cdots, n$$

试求 $\beta_0+\beta_1 x_0$ 的水平为 α 的置信区间.

10. 设 Y 服从线性模型$(Y, X\beta, \sigma^2 I_n)$，且设矩阵 X 分解为 $k+1$ 块，

$$X=(X_0, X_1, \cdots, X_k)$$

将 β 对应地分解为

$$\beta=\begin{pmatrix}\beta_0\\ \beta_1\\ \vdots\\ \beta_k\end{pmatrix}$$

如果 $X_i^\tau X_j=0$，对 $i, j=1, \cdots, n, i\neq j$，证明:

(1) $\hat{\beta}_L=(X^\tau X)^{-1}X^\tau Y=\begin{pmatrix}\hat{\beta}_0\\ \vdots\\ \hat{\beta}_k\end{pmatrix}$

其中 $\hat{\beta}_i=(X_i^\tau X_i)^{-1}X_i^\tau Y$;

(2) $R_0^2=Y^\tau Y-\sum\limits_{i=0}^{k}\hat{\beta}_i^\tau X_i^\tau Y$

11. 设 Y 服从线性模型 $(Y, X\beta, \sigma^2 V_n)$，$X$ 是 $n\times m(n>m)$满列秩阵，且 V 是已知的 $n\times n$ 满秩阵. 试求 β 的最小二乘估计 $\hat{\beta}_L$ 和 σ^2 的无偏估计，并证明 $\hat{\beta}_L$ 是 β 的最小方差线性无偏估计.

12. 设(X, Y) 服从二元正态分布，则从同一批观察数据获得的 Y 关于 X 的回归直线和 X 关于 Y 的回归直线的估计是否是同一条直线?

13. 设(X, Y)服从三项分布

$$f(x, y)=\frac{m!}{x!y!(m-x-y)!}p^x q^y(1-p-q)^{m-x-y}$$

试求 y 关于 x 的回归曲线.

14. 下面记录了一批婴儿出生体重与6个月时体重的数据，试估计它们之间的相关系数和回归直线。

人数		6个月时体重(Kg)									合计
		5.5—	6.0—	6.5—	7.0—	7.5—	8.0—	8.5—	9.0—	9.5—	
出身体重(Kg)	2.50—		1	1	1	2					5
	2.75—		2	4	5	2	2	1			16
	3.00—		2	2	4	8	4	2			22
	3.25—	1		2	6	2	2	2		1	16
	3.50—			1		4	2	3	1	1	12
	3.75—				1		4	1	1	1	8
	4.00—							1	1		2
合计		1	5	10	17	18	14	10	3	3	81

15. 下面给出了某号钢含碳量与机械性能的试验测量结果，试估计抗拉强度关于含碳量；屈服点关于含碳量的直线回归。

编号	含碳量(x)(%)	抗拉(Y_1)(Kg/mm²)	屈服点(Y_2)(Kg/mm²)	编号	含碳量(x)(%)	抗拉(Y_1)(Kg/mm²)	屈服点(Y_2)(Kg/mm²)
1	0.13	41.25	30	16	0.20	47	32.5
2	0.15	44.5	31.75	17	0.17	46.75	32
3	0.16	45	32	18	0.19	44.5	31
4	0.15	44	31.25	19	0.17	45	32
5	0.17	45.5	31.25	20	0.20	47	32.5
6	0.14	44	30.25	21	0.19	46	34.5
7	0.16	43.5	29	22	0.22	47.5	32
8	0.17	46	32.5	23	0.21	49	35.5
9	0.16	46.75	33.75	24	0.19	45.5	35.5
10	0.14	44.75	30.25	25	0.20	46	33
11	0.17	44.5	30	26	0.21	42	33.5
12	0.16	46.5	31.25	27	0.22	47	31.5
13	0.16	44.25	30.25	28	0.19	50	34.5
14	0.17	47	32	29	0.20	49	35.5
15	0.18	50	36	30	0.24	51.5	34

16．下面给出了小白鼠在接种三种不同菌型伤寒杆菌后的存活日数，试问三种菌型的平均存活日数有否显著差异($\alpha=0.05$)？

菌型	接种后存活日数
I 型	2，4，3，2，4，7，7，2，5，4
II 型	5，6，8，5，10，7，12，6，6
III 型	7，11，6，6，7，9，5，10，6，3，10

17．某医院应用克矽平治疗矽肺，治疗前、中、后期患者血液中粘蛋白含量(mg%)观察结果如下：

患者编号	治疗前	治疗中	治疗后
1	6.5	4.5	3.5
2	7.3	4.4	3.6
3	7.3	5.9	3.7
4	3.0	3.6	2.6
5	7.3	5.5	4.3
6	5.6	4.5	3.7
7	7.3	5.2	5.0

试问用克矽平治疗矽肺对降低血液中粘蛋白含量是否有作用($\alpha=0.05$)？

18．抽查某地区三所小学五年级男学生的身高，得数据如下：

小学	身高数据（厘米）					
第一小学	128.1，	134.1，	133.1，	138.9，	140.8，	127.4
第二小学	150.3，	147.9，	136.8，	126.0，	150.7，	155.8
第三小学	140.6，	143.1，	144.5，	143.7，	148.5，	146.4

试问该地区这三所小学五年级男学生的平均身高有否显著差别($\alpha=0.05$)？

19．四种大白鼠经不同剂量雌激素注射后的子宫重量(克)如下表：

鼠种	雌激素剂量 (mg/100g)		
	0.2	0.4	0.8
甲	106	116	145
乙	42	68	115
丙	70	111	133
丁	42	63	87

试检验：(1) 鼠种的影响是否显著？

(2) 剂量差异的影响是否显著($\alpha=0.05$)？

20．下面记录了某地四个生产队在1956--1959年的小麦平均亩产量：

年　份	生产队产量			
	1	2	3	4
1956	146	200	148	151
1957	258	303	282	290
1958	415	461	431	413
1959	454	452	453	415

试检验：(1) 各生产队间的差异是否显著？

(2) 逐年产量的增长是否显著($\alpha=0.05$)？

21．下面记录了三位操作工分别在四台不同机器上操作三天的日产量：

机　器	操作工		
	甲	乙	丙
M_1	15, 15, 17	19, [illegible], 16	16, 18, 21
M_2	17, 17, 17	15, 15, 15	19, 22, 22
M_3	15, 17, 16	18, 17, 16	18, 18, 18
M_4	18, 20, 22	15, 16, 17	17, 17, 17

试检验：(1) 操作工之间的差异是否显著？

(2) 机器之间的差别是否显著？

(3) 交互影响是否显著($\alpha=0.05$)？

22．试运用广义似然比方法推导方差分析模型中统计假设的检验方法。

目　　录

第三册　随机过程

第一章　引论……1

§1. 随机过程的概念及其作用……1

§2. 有限维分布族与柯尔莫哥洛夫(Колмогоров)定理……5

§3. 随机过程的基本类型……10

第二章　马尔可夫过程……34

§1. 马尔可夫链……34

§2. 纯不连续马尔可夫过程……60

§3. 扩散过程……89

习题……108

第三章　二阶矩过程和随机分析……115

§1. 预备知识……115

§2. 随机分析……122

§3. 正态过程……140

§4. 伊藤随机积分和伊藤随机微分方程……149

习题……170

第四章　平稳过程……174

§1. 平稳过程和协方差函数……174

§2. 平稳过程和协方差函数的谱分解……187

§3. 线性系统中的平稳过程……207

§4. 自回归滑动和过程……228

习题……249

第五章　时间序列分析……253

§1. 时间序列的预测和滤波……254

§2. 线性模型的均值估计……282

§3. 余差的协方差和谱的估计……301

§4. 自回归模型拟合……332
习题……350
第六章 随机模拟……352
§1. 基本概念……352
§2. 随机变量的模拟……358
§3. 随机过程的模拟……377
§4. 一些应用……382

第一章 引 论

§1. 随机过程的概念及其作用

在前两册中，我们学习了随机变量、随机向量，即多维随机变量的知识，那儿主要涉及一个或有限个随机变量. 在极限定理中，我们涉及了无穷多个随机变量，但他们之间是相互独立的. 在这一部份中，我们将研究一族无穷多个、相互有关的随机变量，记为 $\{X(t), t\in T\}$，其中 T 是一个无穷集合. 最常见的有：(1) $T_1=\{0, 1, 2, 3, \cdots\}$，(2) $T_2=\{\cdots, -3, -2, -1, 0, 1, 2, 3, \cdots\}$，(3) $T_3=[a, b]$，其中 a, b 可以是 $\pm\infty$. 由于 t 一般表示时间，通常就称它们为随机过程. 所以，**随机过程**就是一族无穷多个随机变量 $\{X(t), t\in T\}$，也可称它为**随机函数**，其中的 T 又称为**参数集**. 今后除特别声明外，参数集 T 总假定是上述三种形式之一. 当 T 为 (1)，(2) 两种情形时，也称为**随机序列**.

随机过程的理论产生于本世纪的初期，是由于物理学、生物学，通讯与控制，管理科学等方面的需要而逐步发展起来的. 下面我们简要地说明一下，在上述诸领域中是怎样提出随机过程的问题，以及随机过程理论的研究对这些领域的发展又有些什么意义.

一、统计物理

随机过程理论的许多部份的发展是与研究物理系统中的噪声密切有关的. 反过来，随机过程理论的发展又为诸如布朗运动，热噪声等物理现象提供了数学模型，更进一步为统计物理奠定了数学基础.

布朗运动 英国植物学家布朗注意到，漂浮在液面上的微小

粒子，不断地进行着杂乱无章的运动．后来这种运动就称为**布朗运动**．从统计物理学的观点来看，粒子的这种运动是由于受到大量随机的、相互独立的分子碰撞的结果．如用$(X(t),\ Y(t))$表示时刻t时粒子的位置，则由于运动是杂乱无章的，$X(t)$和$Y(t)$都是随机变量，于是描写粒子位置的$(X(t),Y(t))$就是一个随机过程．

热噪声 考虑电子网络中的一个电阻．由于电阻中自由电子的随机运动，导至电阻两端的电压$X(t)$有一个随机的起伏．这一起伏的电压就称为**热噪声**．它也是依赖于时间t的一族无穷多个随机变量，于是它是一个随机过程．

散弹噪声 考虑一个真空二极管，由于自热阴极发射出来的电子是不稳定的，因而从阳极到阴极的电流就有随机的起伏．这一随机起伏的电流$X(t)$就称为**散弹噪声**．它当然也是一个随机过程．

二、群体生长的随机模型

一个群体(例如可以是某公社饲养的猪的全体，一个宇宙射线粒子引起的裂变的原子全体等等)的大小和组成通常是有起伏的．如用$X(t)$表示时刻t时群体的大小，则$\{X(t),\ t\in T\}$是一个随机过程．

随机过程理论为与群体生长有关的许多生物和物理现象的研究提供了数学模型．它们包括：灭种问题；数量遗传学；竞争现象；传染病扩散；癌细胞扩散等等．

三、管理科学

排队 顾客来到服务站要求服务．当服务站中的服务员都忙碌时，即当服务员都正在为别的顾客服务时，来到的顾客就要排队等待服务．由于顾客的来到一般是随机的，每个顾客所需的服务时间也是随机的，因此如用$X(t)$表示时刻t时的队长，用$Y(t)$表

示时刻 t 时来到的顾客所需的等待时间等，这些都是随机的．因而 $\{X(t), t\in T\}$ 等都是随机过程．

由于“顾客”与“服务员”的概念是相当的广泛，例如，顾客可以是电话的“呼唤”，“服务员”则是“电话交换总机”；“顾客”可以是“待修复的机器”，“服务员”则是“修理工人”；“顾客”也可以是“待加工的程序”，“服务员”则是“中央处理器”等等．因此关于排队的理论，即排队论有着广泛的应用，而排队论也正是随机过程理论的一个重要分支．历史上，排队现象（特别是电话问题）中的随机过程，如上述的队长，等待时间等也是最早被研究的随机过程之一．

存储控制　在存储控制中的两个重要问题是：（1）何时要开始订货；（2）每次订货的量以多大为宜．这两个问题的回答，当然依赖于货物的销售速度和每次从订货到得到货物所需的时间．而这两者显然都不是确定的．如果用 $X(t)$ 表示在 $[0, t]$ 中的销售量，$Y(t)$ 表示时刻 t 时发出的订货到真正得到所订货物需要的时间等，则由于各种复杂的因素，$\{X(t), t\in T\}$ 等都是随机过程．由此可见，有关随机过程的研究将为进行最优的存储控制提供依据．

四、预测和控制

在与预测和控制有关的问题中，出现大量的随机过程，它们正是随机过程理论发展的一个重要推动力．

预测　预测问题一般可表述为：如何根据对随机过程 $X(t)$ 在时间间隔 $s-L\leqslant t\leqslant s(L>0)$ 上的观察来估计以后某时刻 $s+\tau$ 时的值 $X(t+\tau), \tau>0$．

如果 $X(t)$ 表示时刻 t 时的某处的气温，那便是天气预报问题；如果 $X(t)$ 表示时刻 t 时的某处的水位，那便是水位预报；$X(t)$ 也可以表示时刻 t 时敌机的高度，那么上述预测对于火炮的控制便有很大的意义．

参数估计 考虑一个雷达站，它在 $t=0$ 开始发出一个信号，遇到障碍物后又反射回来，在 $t=\tau$ 时雷达站开始收到反射信号。若时刻 t 发出的信号为 $s(t)$，则此时收到的反射信号可表示为 $A\cdot s(t-\tau)$，其中 A 反映信号经反射后的能量损失，τ 反映雷达站与障碍物之间的距离。由于各种随机干扰的存在，雷达站实际接收到的是 $X(t)=A\cdot s(t-\tau)+N(t)$，其中 $N(t)$ 反映各种随机干扰的效应，它是随机过程。雷达信号处理中的一个十分重要的问题就是如何从被干扰(也称为噪声)模糊了的 $X(t)$ 中，把参数 τ 尽可能精确的估计出来，估计出 τ，也就确定了障碍物到雷达站的距离 $d=\frac{c}{2}\cdot\tau$，其中 c 是光速。显然，这一问题的解决有赖于对反映干扰的 $N(t)$ 这一随机过程的特性的研究。

上述问题是有代表性的，类似的问题在涉及有干扰存在时的动态测量中都是存在的。

五、时间序列分析

在本书的第一册中，我们学习了随机变量及其描述(分布，数字特征)等。在第二册中，我们学习了如何从随机变量的观察中，对随机变量的分布，数字特征等进行估计和推断。类似地，在随机过程的理论中，我们也要研究如何从对随机过程的观察中，对随机过程的统计特性进行估计和推断的方法，这也就是随机过程的统计理论。

对随机过程 $X(t)$ 进行一次观察，得到的便是一个确定的普通意义下的函数 $X(t)$，它称为相应随机过程 $X(t)$ 的一个**"现实"**或**"样本函数"**。在时间参数为离散的场合，就是一个序列，通常称为**"时间序列"**，"现实"、"样本函数"、"时间序列"等术语有时也相互通用。

时间序列的例子是十分广泛的：(1) 某地历年来的月平均气

温；(2) 某地下水水位观察井历年来的旬地下水水位；(3) 某一品种鸡的平均日产蛋个数；(4) 国际市场上历年小麦与大米价格比等等.

时间序列分析的一个重要课题是如何根据对时间序列的研究分析，建立起一个合适的随机过程模型，以用于预测或控制等目的.

§2. 有限维分布族与柯尔莫哥洛夫(Колмогоров)定理

设$\{X(t), t\in T\}$是一个随机过程，为了描述它的统计特性，我们自然要知道每个 $t\in T$ 的$X(t)$的分布函数

$$F(t;x)\triangleq P\{X(t)<x\},\ t\in T$$

我们称$F(t;x)$为随机过程$\{X(t), t\in T\}$的**一维分布**，与此有关的，我们引入(当然假定下式右端存在)：

$$m(t)\triangleq E\{X(t)\}$$

$$D(t)\triangleq E\{(X(t)-m(t))^2\}$$

分别称它们为随机过程$\{X(t), t\in T\}$的**均值函数和方差函数**. 显然只有一维分布，还不足以完全描述随机过程. 为了研究随机过程中任意两个随机变量之间的关系，我们再引入

$$F(t_1, t_2; x_1, x_2)\triangleq P\{X(t_1)<x_1, X(t_2)<x_2\},\ t_1, t_2\in T$$

称它们为随机过程$\{X(t), t\in T\}$的**二维分布**，与此有关的，我们引入(当然假定下式右边存在)

$$\begin{aligned}\Gamma(s, t)&=\text{cov}\{X(s), X(t)\}\\&=E\{(X(s)-m(s))(X(t)-m(t))\}\end{aligned}$$

它称为随机过程$\{X(t), t\in T\}$的**协方差函数**. 显然有：

$$D(t)=\Gamma(t, t)$$

一般地，对任意有限个 $t_1, t_2, \cdots, t_n\in T$，引入

$$F(t_1, t_2, \cdots, t_n; x_1, x_2, \cdots, x_n)$$
$$\triangleq P\{X(t_1)<x_1, X(t_2)<x_2, \cdots, X(t_n)<x_n\}$$

并称它们为随机过程$\{X(t), t\in T\}$的 n **维分布**.

随机过程 $\{X(t), t\in T\}$ 的一维分布, 二维分布, …, n 维分布等等, 其全体$\{F(t_1, t_2, \cdots, t_n; x_1, \cdots, x_n), t_1, \cdots, t_n\in T, n\geqslant 1\}$称为随机过程$\{X(t), t\in t\}$的**有限维分布族**. 如上所述可见, 如果知道了随机过程$\{X(t), t\in T\}$的 n 维分布全体, 则对随机过程$\{X(t), t\in T\}$中任意 n 个随机变量的联合分布也就完全知道了, 如果我们知道了随机过程的有限维分布族, 便知道了这一随机过程中任意有限个随机变量的联合分布, 也就可以完全确定它们之间的相互关系等等.

由多维分布函数的性质和上述定义, 容易看出, 一个随机过程的有限维分布族具有如下两个性质.

(1) **对称性** 对$(1, 2, \cdots, n)$的任一排列$(j_1, j_2, \cdots, j_n)$, 有

$$F(t_{j_1}, t_{j_2}, \cdots, t_{j_n}; x_{j_1}, x_{j_2}, \cdots, x_{j_n})$$
$$=F(t_1, t_2, \cdots, t_n; x_1, x_2, \cdots, x_n)$$

(2) **相容性** 对 $m<n$, 有

$$F(t_1, t_2, \cdots, t_m, t_{m+1}, \cdots, t_n; x_1, x_2, \cdots, x_m, \infty, \cdots, \infty)$$
$$=F(t_1, t_2, \cdots, t_m; x_1, x_2, \cdots, x_m)$$

引入特征函数

$$\phi(t_1, t_2, \cdots, t_n; \theta_1, \theta_2, \cdots, \theta_n)$$
$$=E\{e^{i(\theta_1 X(t_1)+\theta_2 X(t_2)+\cdots+\theta_n X(t_n))}\}$$
$$=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} e^{i(\theta_1 x_1+\cdots+\theta_n x_n)} F(t_1, \cdots, t_n; dx_1, \cdots, dx_n)$$

称$\{\phi(t_1, \cdots, t_n; \theta_1, \cdots, \theta_n), n\geqslant 1, t_1, \cdots, t_n\in T\}$为随机过程$\{X(t), t\in T\}$的**有限维特征函数族**. 由特征函数唯一性定理(第一册§4.4)可知有限维分布族与有限维特征函数族是相互唯一决定的, 有限

维分布族的对称性和相容性对有限维特征函数族表现为:

(1) **对称性** 对$(1,2,\cdots,n)$的任一排列$(j_1,j_2,\cdots,j_n)$,有

$$\phi(t_{j_1},t_{j_2},\cdots,t_{j_n};\theta_{j_1},\theta_{j_2},\cdots,\theta_{j_n})$$
$$=\phi(t_1,t_2,\cdots,t_n;\theta_1,\theta_2,\cdots,\theta_n)$$

(2) **相容性** 对$m<n$,有

$$\phi(t_1,t_2,\cdots,t_m,t_{m+1},\cdots,t_n;\theta_1,\theta_2,\cdots,\theta_m,0,\cdots,0)$$
$$=\phi(t_1,t_2,\cdots,t_m;\theta_1,\theta_2,\cdots,\theta_m)$$

苏联数学家柯尔莫哥洛夫在1931年证明了如下的基本定理:

定理(柯尔莫哥洛夫) 设分布函数族$\{F(t_1,t_2,\cdots,t_n;x_1,x_2,\cdots,x_n),n\geqslant1,t_1,t_2,\cdots,t_n\in T\}$,满足上述对称性和相容性,则必存在一个随机过程$\{X(t),t\in T\}$,使$\{F(t_1,t_2,\cdots,t_n;x_1,x_2,\cdots,x_n),n\geqslant1,t_1,t_2,\cdots,t_n\in T\}$恰好是$X(t)$的有限维分布族,即

$$F(t_1,t_2,\cdots,t_n;x_1,x_2,\cdots,x_n)$$
$$=P\{X(t_1)<x_1,X(t_2)<x_2,\cdots,X(t_n)<x_n\}$$

定理的证明比较复杂,这儿从略①.

为了帮助理解,我们来看几个简单的随机过程的例子.

[例1] $X(t)=X_0+V\cdot t,a\leqslant t\leqslant b$,其中$X_0$和$V$是相互独立的正态$N(0,1)$分布的随机变量.

由于X_0和V都是正态分布,所以$X(t)$也是正态分布,并且$X(t_1),X(t_2),\cdots,X(t_n)$也是$n$维正态分布.对于正态分布,只要知道它们的一阶矩和二阶矩就完全确定了它们的分布.

对于我们现在这个过程,一阶矩和二阶矩是容易求得的:

$$m(t)=E\{X(t)\}=E\{X_0+V\cdot t\}$$
$$=E\{X_0\}+tE\{V\}=0$$
$$\Gamma(s,t)=E\{X(s)X(t)\}$$

---

① 参见王梓坤,随机过程论,科学出版社,1965,第8页.

$$=E\{(X_0+V\cdot s)(X_0+V\cdot t)\}$$
$$=E\{X_0^2\}+E\{V^2\}s\cdot t=1+st$$

根据前面所述，现在已经很容易写出 $X(t)$ 的有限维分布了．例如

$$F(t;x)=\frac{1}{\sqrt{2\pi(1+t^2)}}\int_{-\infty}^{x}e^{-\frac{u^2}{2(1+t^2)}}du$$

［例 2］ $X(t)=A\cdot\cos\theta t+B\cdot\sin\theta t,\ 0\leqslant t\leqslant 1$，其中 A,B 是相互独立的正态 $N(0,\sigma^2)$ 分布的随机变量，θ 是一个实常数．

由于这里的 $X(t)$ 是正态随机变量 A，B 的线性组合，所以 $X(t)$ 也是正态分布的，同样，$X(t_1)$，$X(t_2)$，…，$X(t_n)$ 是 n 维正态分布的，这时要知道 $X(t)$ 的有限维分布，只要知道它的一阶矩和二阶矩，即它的均值函数和协方差函数就可以了．

我们有：

$$m(t)=E\{X(t)\}$$
$$=E\{A\cdot\cos\theta t+B\cdot\sin\theta t\}=0$$
$$\Gamma(s,t)=E\{X(s)\overline{X(t)}\}$$
$$=E\{(A\cdot\cos\theta s+B\cdot\sin\theta s)\cdot(A\cdot\cos\theta t+B\cdot\sin\theta t)\}$$
$$=\sigma^2(\cos\theta s\cdot\cos\theta t+\sin\theta s\cdot\sin\theta t)$$
$$=\sigma^2\cdot\cos\theta(s-t)$$

［例 3］ $Z(t)=\sum_{k=1}^{N}A_k e^{i\theta_k t},\ 0\leqslant t\leqslant 1$，其中 $A_k,\ 1\leqslant k\leqslant N$ 是相互独立的 $N(0,\sigma_k^2)$ 分布的随机变量，$i=\sqrt{-1}$ 是虚数单位．

由欧拉公式 $e^{ix}=\cos x+i\sin x$ 可知

$$Z(t)=\sum_{k=1}^{N}A_k\cdot\cos\theta_k t+i\cdot\sum_{k=1}^{N}A_k\cdot\sin\theta_k t$$

因此 $Z(t)$ 一般是复的．对于复的随机变量 $Z=X+iY$，其中 X,Y 都是实的随机变量，我们称 X,Y 的联合分布为 Z 的分布，称 $EX+$

iEY(如果EX和EY都存在)为Z的均值，记为EZ. 对于两个复的随机变量$Z_1=X_1+iY_1$, $Z_2=X_2+iY_2$定义它们的协方差为:

$$\operatorname{cov}(Z_1, Z_2)=E\{(Z_1-EZ_1)\cdot(\overline{Z_2-EZ_2})\}$$

其中"——"表示"共轭". 这是与实的随机变量时的定义形式上不同的地方，但这一定义应用于实的随机变量时，由于这时共轭与否是一样的，所以结果是一样的.

由于 $Z(t)$ 的实部 $X(t)=\sum_{k=1}^{N}A_k\cos\theta_k t$ 和虚部 $Y(t)=\sum_{k=1}^{N}A_k\sin\theta_k t$ 都是正态随机变量 A_k, $1\leqslant k\leqslant N$, 的线性组合，所以它们也都是正态的，因此也只要求出它们的均值函数与协方差函数就可以了.

我们有:

$$\begin{aligned}
m(t)&=E\{Z(t)\}\\
&=E\{X(t)\}+iE\{Y(t)\}=0\\
\Gamma(s,t)&=E\{Z(s)\overline{Z(t)}\}\\
&=E\left\{\left(\sum_{l=1}^{N}A_le^{i\theta_l s}\right)\cdot\overline{\left(\sum_{k=1}^{N}A_ke^{i\theta_k t}\right)}\right\}\\
&=\sum_{l,k=1}^{N}E\{A_lA_k\}e^{i(\theta_l\cdot s-\theta_k\cdot t)}\\
&=\sum_{k=1}^{N}\sigma_k^2e^{i\theta_k(s-t)}
\end{aligned}$$

类似地可以求得$Z(t)$的实部$X(t)$和虚部$Y(t)$的均值函数和协方差函数.

$$\begin{aligned}
m_x(t)&=E\{X(t)\}=0\\
\Gamma_x(s,t)&=E\{X(s)X(t)\}\\
&=\sum_{k=1}^{N}\sigma_k^2\cdot\cos(\theta_k s)\cos(\theta_k t)
\end{aligned}$$

$$m_y(t)=0$$

$$\Gamma_y(s,t)=E\{Y(s)Y(t)\}$$

$$=\sum_{k=1}^{N}\sigma_k^2\cdot\sin(\theta_k s)\sin(\theta_k t)$$

从上面的例子可以看出，如果随机过程$\{X(t),\ t\in T\}$的有限维分布都是正态分布，那么要知道它的有限维分布族，只要知道它的均值函数 $m(t)$和协方差函数 $\Gamma(s,t)$就足够了。

§3. 随机过程的基本类型

一般的随机过程是相当复杂和难以研究的，已有的结果也比较少．所以在本书中，主要研究两类最重要的随机过程：马尔可夫过程和二阶矩过程．

一、二阶矩过程

如果随机过程$\{X(t),\ t\in T\}$，对每个 $t\in T$，$X(t)$的均值和方差都存在，则称它为**二阶矩过程**．特别，如果$\{X(t),\ t\in T\}$的有限维分布都是正态分布，则称为**正态过程**，正态过程是二阶矩过程的一个最重要的子类．

由上述定义立即可知，上节的三个随机过程的例子，都是属于二阶矩过程的，而且还都是正态过程．

由定义可知，对于二阶矩过程$\{X(t),\ t\in T\}$，它的均值函数 $m(t)=E\{X(t)\}$ 是存在的．因此，如从 $X(t)$中减去 $m(t)$，即代替$\{X(t),\ t\in T\}$而考虑 $\widetilde{X}(t)=X(t)-m(t)$，则有：

$$m_{\widetilde{X}}(t)=E\{\widetilde{X}(t)\}=0$$

因此，今后一般都假定二阶矩过程的均值函数为零．这时，协方差函数简化为：

$$\Gamma(s,t)=E\{X(s)\overline{X(t)}\}$$

应用许瓦兹(Schwartz)不等式，可以证明，二阶矩过程的协方差函数总是存在的. 事实上，由许瓦兹不等式有

$$(E\{|X(s)\overline{X(t)}|\})^2 \leqslant E\{|X(s)|^2\}\cdot E\{|X(t)|^2\}$$

而根据定义，上式右边是有限的.

下面给出协方差函数的两个性质.

定理 1 二阶矩过程的协方差函数 $\Gamma(s, t)$ 是埃密特的，即

$$\Gamma(s, t)=\overline{\Gamma(t, s)}, \qquad s, t \in T$$

如果过程是实的，即 $X(t)$ 是实的随机变量，则 $\Gamma(s, t)$ 是对称的.

[证明] 我们有

$$\begin{aligned}\overline{\Gamma(t, s)} &= \overline{E\{X(t)\overline{X(s)}\}} \\ &= E\{\overline{X(t)}X(s)\} \\ &= E\{X(s)\overline{X(t)}\}=\Gamma(s, t)\end{aligned}$$

定理 2 二阶矩过程的协方差函数 $\Gamma(s, t)$ 是非负定的，即对任意有限个 $t_1, t_2, \cdots, t_n \in T$ 和任意的普通函数 $\theta(t)$, $t \in T$, 有

$$\sum_{k=1}^{n}\sum_{j=1}^{n}\Gamma(t_k, t_j)\theta(t_k)\overline{\theta(t_j)} \geqslant 0$$

[证明] 我们有

$$\begin{aligned}&\sum_{k=1}^{n}\sum_{j=1}^{n}\Gamma(t_k, t_j)\theta(t_k)\overline{\theta(t_j)} \\ &= \sum_{k=1}^{n}\sum_{j=1}^{n}E\{X(t_k)\overline{X(t_j)}\}\theta(t_k)\overline{\theta(t_j)} \\ &= E\left\{\left(\sum_{k=1}^{n}X(t_k)\theta(t_k)\right)\cdot\overline{\left(\sum_{j=1}^{n}X(t_j)\theta(t_j)\right)}\right\} \\ &= E\left\{\left|\sum_{k=1}^{n}X(t_k)\theta(t_k)\right|^2\right\} \geqslant 0\end{aligned}$$

顺便指出，由非负定性可以推出埃密特性. 因此，定理 2 包含了定理 1.

事实上，先令$n=1$，和$\theta(t_1)=1$，则由非负定性，可得

$$\Gamma(t_1, t_1)\geqslant 0$$

再令$n=2$，和$\theta(t_1)=1$，$\theta(t_2)=i(i=\sqrt{-1})$，则由非负定性，可得

$$\Gamma(t_1, t_1)-\Gamma(t_1, t_2)i+\Gamma(t_2, t_1)i$$
$$+\Gamma(t_2, t_2)\geqslant 0$$

又因$\Gamma(t_1, t_1)\geqslant 0$，$\Gamma(t_2, t_2)\geqslant 0$都是实的，所以

$$-\Gamma(t_1, t_2)i+\Gamma(t_2, t_1)i$$

应是实的．于是$\Gamma(t_1, t_2)$的实部应与$\Gamma(t_2, t_1)$的实部相等．类似地，令$n=2$，$\theta(t_1)=\theta(t_2)=1$，用同样的方法可知

$$\Gamma(t_1, t_2)+\Gamma(t_2, t_1)$$

也应是实的．于是$\Gamma(t_1, t_2)$的虚部应与$\Gamma(t_2, t_1)$的虚部有相同的绝对值，但符号相反．综合上述两点，即得

$$\Gamma(t_1, t_2)=\overline{\Gamma(t_2, t_1)}$$

下面是定理2的逆．

定理3　在$T\times T$上给定函数$\Gamma(s, t)$，如果它是非负定的，则一定存在一个二阶矩过程（事实上还是正态过程）$\{X(t), t\in T\}$，使$\Gamma(s, t)$正好就是$\{X(t), t\in T\}$的协方差函数．如果$\Gamma(s, t)$是实的，则相应的随机过程也是实的．

［证明］　先考虑$\Gamma(s, t)$是实的情形．此时令$\theta(t)$也是实的，则按定理的条件，对任意有限个$t_1, t_2, \cdots, t_n\in T$，

$$\sum_{k=1}^{n}\sum_{j=1}^{n}\Gamma(t_k, t_j)\theta(t_k)\theta(t_j)\geqslant 0$$

于是按多维正态分布定义（第一册定理5.4.1），下列函数

$$\exp\left(-\frac{1}{2}\sum_{k=1}^{n}\sum_{j=1}^{n}\Gamma(t_k, t_j)\theta(t_k)\theta(t_j)\right)$$

是均值都为零，协方差矩阵为

$$\begin{bmatrix} \Gamma(t_1, t_1) & \Gamma(t_1, t_2) & \cdots & \Gamma(t_1, t_n) \\ \Gamma(t_2, t_1) & \Gamma(t_2, t_2) & \cdots & \Gamma(t_2, t_n) \\ \cdots & \cdots & \cdots & \cdots \\ \Gamma(t_n, t_1) & \Gamma(t_n, t_2) & \cdots & \Gamma(t_n, t_n) \end{bmatrix}$$

的 n 维正态分布的特征函数. 记它相应的 n 维正态分布函数为:

$$F(t_1, t_2, \cdots, t_n; x_1, x_2, \cdots, x_n)$$

容易看出, 其全体 $\{F(t_1, t_2, \cdots, t_n; x_1, x_2, \cdots, x_n), n\geqslant 1, t_1, t_2, \cdots, t_n\in T\}$ 是满足对称性和相容性条件的分布函数族, 于是根据上节的柯尔莫哥洛夫定理可知, 存在一个随机过程 $\{X(t), t\in T\}$, 它以如上定义的多维正态分布函数族

$$\{F(t_1, t_2, \cdots, t_n; x_1, x_2, \cdots, x_n), n\geqslant 1, t_1, t_2, \cdots, t_n\in T\}$$

作为自己的有限维分布族. 特别有:

$$E\{X(t)\}=0$$

$$E\{X(s)X(t)\}=\Gamma(s, t), \qquad s, t\in T$$

现在来考虑 $\Gamma(s, t)$ 是复的情形. 令

$$\theta(t)=\alpha(t)-i\beta(t)$$

$$\Gamma(s, t)=A(s, t)+iB(s, t), \qquad s, t\in T$$

其中 $\alpha(t), \beta(t), A(s, t), B(s, t)$ 都是实值函数. 按定理条件, 有

$$\begin{aligned} 0\leqslant Q&=\sum_{k=1}^{n}\sum_{j=1}^{n}\Gamma(t_k, t_j)\theta(t_k)\overline{\theta(t_j)} \\ &=\sum_{k=1}^{n}\sum_{j=1}^{n}(A(t_k, t_j)+iB(t_k, t_j)) \\ &\quad\cdot(\alpha(t_k)-i\beta(t_k))(\alpha(t_j)+i\beta(t_j)) \end{aligned}$$

由于 $Q\geqslant 0$, 因此一定是实的, 所以上式的虚部一定是零, 因此有

$$\begin{aligned} Q=&\sum_{k=1}^{n}\sum_{j=1}^{n}A(t_k, t_j)(\alpha(t_k)\alpha(t_j)+\beta(t_k)\beta(t_j)) \\ &-\sum_{k=1}^{n}\sum_{j=1}^{n}B(t_k, t_j)(\alpha(t_k)\beta(t_j)-\alpha(t_j)\beta(t_k))\geqslant 0 \end{aligned}$$

于是按多维正态分布的定义(第一册定理 4.5.1)下列函数

$$\exp\left(-\frac{1}{4}Q\right)$$

是均值为零,协方差矩阵为:

$$\begin{bmatrix} \frac{1}{2}A(t_1, t_1), \cdots, \frac{1}{2}A(t_1, t_n), -\frac{1}{2}B(t_1, t_1), \cdots, -\frac{1}{2}B(t_1, t_n) \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \frac{1}{2}A(t_n, t_1), \cdots, \frac{1}{2}A(t_n, t_n), -\frac{1}{2}B(t_n, t_1), \cdots, -\frac{1}{2}B(t_n, t_n) \\ \frac{1}{2}B(t_1, t_1), \cdots, \frac{1}{2}B(t_1, t_n), \quad \frac{1}{2}A(t_1, t_1), \cdots, \quad \frac{1}{2}A(t_1, t_n) \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \frac{1}{2}B(t_n, t_1), \cdots, \frac{1}{2}B(t_n, t_n), \quad \frac{1}{2}A(t_n, t_1), \cdots, \quad \frac{1}{2}A(t_n, t_n) \end{bmatrix}$$

的 $2n$ 维正态分布的特征函数.

仿前可证,存在一个二元正态过程$\{X(t), Y(t), t\in T\}$,它以上述的多维正态分布为自己的有限维分布. 特别有:

$$E\{X(t)\}=E\{Y(t)\}=0$$

$$E\{X(s)X(t)\}=E\{Y(s)Y(t)\}$$

$$=\frac{1}{2}A(s, t)$$

$$E\{X(s)Y(t)\}=-\frac{1}{2}B(s, t)$$

于是,如定义

$$Z(t)=X(t)+iY(t)$$

则有:

$$E\{Z(t)\}=E\{X(t)\}+iE\{Y(t)\}=0$$

$$E\{Z(s)\overline{Z(t)}\}=E\{(X(s)+iY(s))(X(t)-iY(t))\}$$

$$=E\{X(s)X(t)\}+E\{Y(s)Y(t)\}-iE\{X(s)Y(t)\}$$
$$+iE\{X(t)Y(s)\}$$

$$=\frac{1}{2}A(s,t)+\frac{1}{2}A(s,t)+\frac{i}{2}B(s,t)-\frac{i}{2}B(t,s)$$

再由 $\Gamma(s,t)$是埃密特的，可得

$$\begin{aligned}A(s,t)+iB(s,t)&=\Gamma(s,t)\\&=\overline{\Gamma(t,s)}\\&=A(t,s)-iB(t,s)\end{aligned}$$

于是有：

$$A(s,t)=A(t,s)$$

$$B(s,t)=-B(t,s)$$

亦即协方差函数的实部是对称的，而虚部是反对称的．利用后一事实，可得：

$$\begin{aligned}E\{Z(s)\overline{Z(t)}\}&=A(s,t)+\frac{i}{2}B(s,t)+\frac{i}{2}B(s,t)\\&=A(s,t)+iB(s,t)\\&=\Gamma(s,t)\end{aligned}$$

于是定理 3 证毕．

对于二阶矩过程，我们可以由它的分布确定它的均值函数和协方差函数．反过来，由随机过程的均值函数和协方差函数一般并不能完全确定随机过程的全部统计特性．但如果我们的过程又是实正态过程或满足(1)的复正态过程时，那么如上所述，由它的均值函数和协方差函数就可以完全确定这一过程的有限维分布族，从而完全确定了随机过程的统计特性．

在第三章中，我们将进一步研究二阶矩过程，那儿将引进二阶矩过程的连续性，可微性，可积性等等概念，以及有关的运算法则等，这是普通数学分析在随机场合的一种推广．

在二阶矩过程中，研究得最多的是一种称为宽平稳过程的随机过程．一个二阶矩过程$\{X(t),\ t\in T\}$称为**宽平稳的**，或**宽平稳**

过程，如果它的均值函数

$$m(t)=m=\text{常数}$$

而它的协方差函数 $\Gamma(s,t)$ 只与 $s-t$ 有关，而与 s，t 本身无关，亦即如果存在一个一元函数 $B(\tau)$，使

$$\Gamma(s,t)=B(s-t), \quad s,t\in T$$

其中 $B(\tau)$ 称为宽平稳过程的**协方差函数**.

容易看到，上节中的例 1 不是平稳过程，而例 2 与例 3 中的 $Z(t)$ 都是宽平稳过程. 相应的 $B(\tau)$ 是

$$B(\tau)=\sigma^2\cdot\cos\theta\tau$$

和

$$B(\tau)=\sum_{k=1}^{n}\sigma_k^2 e^{i\theta_k\tau}$$

由 $\Gamma(s,t)$ 的埃密特性和非负定性，易知宽平稳过程的协方差函数有下列性质：

(1) 埃密特性：$B(-\tau)=\overline{B(\tau)}$；

(2) 非负定性：即对任意有限个 $t_1, t_2, \cdots, t_n\in T$ 和任意的普通函数 $\theta(t)$，$t\in T$，有

$$\sum_{k=1}^{n}\sum_{j=1}^{n}B(t_k-t_j)\theta(t_k)\overline{\theta(t_j)}\geqslant 0$$

由波赫纳尔-辛钦定理(第一册定理 5.2.9)可知，对于满足上述性质的 $B(\tau)$，若 $B(\tau)$ 在 $\tau=0$ 连续，则存在一个非降的实值函数 $F(\lambda)$，使

$$B(\tau)=\int_{-\infty}^{\infty}e^{i\tau\lambda}dF(\lambda)$$

其中 $F(\lambda)$ 称为宽平稳过程的**谱函数**，上式称为**宽平稳过程协方差函数的谱分解**. 特别如果 $F'(\lambda)=f(\lambda)$ 存在，则上式变为：

$$B(\tau)=\int_{-\infty}^{\infty}e^{i\tau\lambda}f(\lambda)d\lambda$$

其中 $f(\lambda)\geqslant 0$(因为 $F(\lambda)$ 非降)称为宽平稳过程的**谱密度**.

如果过程是实的,则相应的 $B(\tau)$ 也是实的,当 $f(\lambda)$ 存在时,

$$B(\tau)=\overline{B(\tau)}=\int_{-\infty}^{\infty}e^{i\tau\lambda}f(\lambda)d\lambda=\int_{-\infty}^{\infty}e^{-i\tau\lambda}f(\lambda)d\lambda$$

$$=\int_{-\infty}^{\infty}e^{i\tau\lambda}f(-\lambda)d\lambda$$

由于上式对一切 τ 成立,所以 $f(\lambda)=f(-\lambda)$. 若取 $f_1(\lambda)=2f(\lambda)$,则

$$B(\tau)=\int_{-\infty}^{\infty}e^{i\tau\lambda}f(\lambda)d\lambda=\int_{0}^{\infty}\cos\tau\lambda f_1(\lambda)d\lambda$$

而对一般的实过程,若取 $F_1(\lambda)=2F(\lambda)+C$,则其协方差函数谱分解可写为

$$B(\tau)=\int_{0}^{\infty}\cos\tau\lambda dF_1(\lambda)$$

容易看出,上节的例 2 是实的,相应的

$$F_1(\lambda)=\begin{cases}\sigma^2, & \lambda>\theta\\ 0, & \lambda\leqslant\theta\end{cases}$$

此时,$f(\lambda)$ 不存在. 而例 3 是复的,相应的

$$F(\lambda)=\begin{cases}0, & \lambda\leqslant\theta_1\\ \sum\limits_{k=1}^{n}\sigma_k^2, & \theta_n<\lambda\leqslant\theta_{n+1},\quad 1\leqslant n\leqslant N-1\\ \sum\limits_{k=1}^{N}\sigma_k^2, & \lambda>\theta_N\end{cases}$$

此处假定 $\theta_1\leqslant\theta_2\leqslant\cdots\leqslant\theta_N$. 与例 2 一样,也不存在 $f(\lambda)$.

我们在第四章和第五章中将进一步研究有关宽平稳过程的许多有趣的问题.

与宽平稳过程相对应的,有所谓的严平稳过程. 一个随机过程 $\{X(t),t\in T\}$ 称为**严平稳的**,或**严平稳过程**,如果它的有限分布对

时间推移不变，即如果对任意有限个 $t_1, t_2, \cdots, t_n \in T$ 和任意的实数 τ，当 $t_1+\tau, t_2+\tau, \cdots, t_n+\tau \in T$ 时有：

$$
\begin{aligned}
& F(t_1, t_2, \cdots, t_n; x_1, x_2, \cdots, x_n) \\
&= P\{X(t_1)<x_1, \cdots, X(t_n)<x_n\} \\
&= P\{X(t_1+\tau)<x_1, \cdots, X(t_n+\tau)<x_n\} \\
&= F(t_1+\tau, \cdots, t_n+\tau; x_1, \cdots, x_n)
\end{aligned}
$$

容易看出，一般的严平稳过程未必有二阶矩，因而不一定是宽平稳过程，反之，宽平稳过程也不一定是严平稳过程．但对于正态过程而言，宽平稳与严平稳是一致的．

由于在随机过程的理论中，宽平稳过程受到更多的研究，因此常把宽平稳过程简称为平稳过程．

除了平稳过程外，还有一种比较重要的称为正交增量过程的二阶矩过程．二阶矩过程 $\{X(t), t\in T\}$ 称为**正交增量过程**，如果对于任意的 $t_1<t_2\leqslant t_3\leqslant t_4\in T$，有

$$E\{(X(t_2)-X(t_1))\overline{(X(t_4)-X(t_3))}\}=0$$

对于正交增量过程，若 T 为有限区间 $[a, b]$，且规定 $X(a)=0$．此时，如取 $t_1=a, t_2=t_3=s, t_4=t$，其中 $t>s$，则上式变为：

$$E\{X(s)(\overline{X(t)-X(s)})\}=0$$

对于这类正交增量过程，如定义

$$
\begin{aligned}
F(s) &= E\{X(s)\overline{X(s)}\} \\
&= E\{|X(s)|^2\}
\end{aligned}
$$

则当 $a\leqslant s\leqslant t\leqslant b$ 时，容易求得：

$$
\begin{aligned}
\Gamma(s, t) &= E\{X(s)\overline{X(t)}\} \\
&= E\{X(s)[\overline{X(s)+(X(t)-X(s))}]\} \\
&= E\{X(s)\overline{X(s)}\}+E\{X(s)(\overline{X(t)-X(s)})\} \\
&= E\{|X(s)|^2\}=F(s)
\end{aligned}
$$

一般地，其协方差函数可取为

$$\Gamma(s, t)=F(\min(s, t))$$

而当 $t>s$ 时,

$$0 \leqslant E|X(t)-X(s)|^2=\Gamma(t, t)-\Gamma(t, s)-\Gamma(s, t)+\Gamma(t, t)$$
$$=F(t)-F(s)$$

所以 $F(t)$ 还是单调不减的.

关于正交增量过程,我们在第三章中还要讨论到.

二、马尔可夫过程

另一类重要的随机过程是马尔可夫过程. 为了介绍它的概念, 我们先考虑取值为整数(或它的子集)的随机变量序列 $\{X(n), n=0, 1, 2, \cdots\}$. 假若它具有如下性质: 若对任意的非负整数 $t_1<t_2<\cdots<t_r<t$, 只要

$$P\{X(t_1)=i_1, X(t_2)=i_2, \cdots, X(t_r)=i_r\}>0 \tag{2}$$

就有

$$P\{X(t)=j|X(t_1)=i_1, X(t_2)=i_2, \cdots, X(t_r)=i_r\}$$
$$=P\{X(t)=j|X(t_r)=i_r\} \tag{3}$$

就称 $\{X(n)\}$ 具有**马尔可夫性, 或称无后效性**. 如果把 $X(n)$ 的参数 n 看为时间, t_r 作为"现在", 则 $t>t_r$ 就是"将来", 而 $t_1<t_2<\cdots<t_{r-1}(<t_r)$ 就是"过去". 因而上式也可表述为: 在现在 $X(t_r)=i_r$ 已知的条件下, $\{X(n)\}$ "将来"的情况与"过去"的情况是无关的. 也可以这样说, $\{X(n)\}$ 的将来只是通过"现在"与"过去"发生联系, 一旦"现在"已知, "将来"和"过去"就是无关的了.

为了强调 $\{X(n), n=0, 1, 2, \cdots\}$ 的参数是非负整数, 对具有马尔可夫性的这一随机过程也称作是**马尔可夫链**. 又因为 $X(n)$ 的取值为整数(或它的子集, 今后都表以 I), 也称它为**可列(状态的)马尔可夫链**.

对于马尔可夫链, 描述它概率性质最重要的是它在时刻 m 的**一步转移概率** $p_{ij}(m)$:

$$p_{ij}(m)=P\{X(m+1)=j|X(m)=i\},\qquad i,j\in I$$

它表示时刻 m，$X(m)$取值 i 的条件之下于下一时刻转移到 j 的概率. 虽然(3)中的条件概率只有当(2)成立时才有意义，但在讨论转移概率的性质时往往不强调(2)的成立与否. 显然，$p_{ij}(m)$应具有下列性质：

$$p_{ij}(m)\geqslant 0,\quad i,j\in I$$

$$\sum_{j\in I}p_{ij}(m)=\sum_{j\in I}P\{X(m+1)=j|X(m)=i\}=1,\quad i\in I$$

一般还可定义时刻m的 k 步转移概率如下：

$$p_{ij}^{(k)}(m)=P\{X(m+k)=j|X(m)=i\},\qquad i,j\in I,k\geqslant 1$$

同样，按定义有

$$p_{ij}^{(k)}(m)\geqslant 0$$

$$\sum_{j\in I}p_{ij}^{(k)}(m)=\sum_{j\in I}P\{X(m+k)=j|X(m)=i\}=1,\quad i\in I$$

而

$$p_{ij}^{(1)}(m)=p_{ij}(m)$$

通常，我们还规定

$$p_{ij}^{(0)}(m)=\delta_{ij}=\begin{cases}1, & i=j\\0, & i\neq j\end{cases}$$

对于 k 步转移概率，有如下的切普曼(Chapman)-柯尔莫哥洛夫方程，

$$p_{ij}^{(k+l)}(m)=\sum_{s\in I}p_{is}^{(k)}(m)p_{sj}^{(l)}(m+k),\qquad i,j\in I\qquad(4)$$

为了证明上述等式，我们按 $X(m+k)$的取值不同使用全概率公式可得

$$\begin{aligned}p_{ij}^{(k+l)}(m)&=P\{X(m+k+l)=j|X(m)=i\}\\&=P\{X(m+k+l)=j,\ \bigcup_{s\in I}\{X(k+m)=s\}|X(m)=i\}\end{aligned}$$

$$=\sum_{s\in I}P\{X(m+k+l)=j,X(m+k)=s|X(m)=i\}$$

$$=\sum_{s\in I}P\{X(m+k)=s|X(m)=i\}$$

$$\cdot P\{X(m+k+l)=j|X(m+k)=s,X(m)=i\}$$

$$=\sum_{s\in I}P\{X(m+k)=s|X(m)=i\}$$

$$\cdot P\{X(m+k+l)=j|X(m+k)=s\}$$

$$=\sum_{s\in I}p_{is}^{(k)}(m)p_{sj}^{(l)}(m+k)$$

应当强调指出，在上述推导中倒数第二个等式是利用了 $X(n)$ 的马尔可夫性.

在(4)中特别取 $k=1$ 或 $l=1$ 可得

$$p_{ij}^{(l+1)}(m)=\sum_{s\in I}p_{is}(m)p_{sj}^{(l)}(m+1)$$

$$=\sum_{s\in I}p_{is}^{(l)}(m)p_{sj}(m+l),\quad i,j\in I$$

若再令 $l=1$，则可知由 $\{p_{ij}(m),\ i,\ j\in I,\ m\geq 0\}$ 可决定 $\{p_{ij}^{(2)}(m),\ i,j\in I,m\geq 0\}$. 也就是由全部一步转移概率可确定全部两步转移概率，反复应用上述作法，由全部一步转移概率可以决定所有的转移概率. 进而若再知道初始时刻 $X(0)$ 的分布，即**初始分布** $P\{X(0)=i\}=p_i$，那么 $\{X(n)\}$ 的所有有限维分布也就完全知道了. 例如对 $m_1<m_2<\cdots<m_k$，

$$P\{X(m_1)=i_1,X(m_2)=i_2,\cdots,X(m_k)=i_k\}$$

$$=\sum_{i\in I}P\{X(0)=i\}\cdot P\{X(m_1)=i_1|X(0)=i\}$$

$$\cdot P\{X(m_2)=i_2|X(m_1)=i_1,X(0)=i\}\cdots$$

$$\cdot P\{X(m_k)=i_k|X(m_{k-1})=i_{k-1},\cdots,X(m_1)=i_1,X(0)=i\}$$

$$
\begin{aligned}
&=\sum_{i\in I}P\{X(0)=i\}\cdot P\{X(m_1)=i_1|X(0)=i\}\\
&\qquad\cdot P\{X(m_2)=i_2|X(m_1)=i_1\}\cdots\\
&\qquad\cdot P\{X(m_k)=i_k|X(m_{k-1})=i_{k-1}\}\\
&=\sum_{i\in I}p_ip_{ii_1}^{(m_1)}(0)p_{i_1i_2}^{(m_2-m_1)}(m_1)\cdots p_{i_{k-1}i_k}^{(m_k-m_{k-1})}(m_{k-1})
\end{aligned}
$$

在上述推导中我们也利用了$X(n)$的马尔可夫性. 所以, 虽然转移概率$p_{ij}(m)$只描述了$X(m)$与$X(m+1)$之间的统计联系, p_i只描写了$X(0)$的分布. 但是利用了$X(n)$的马尔可夫性, 可以由它们描写出$X(n)$的一切有限维分布律, 因而也可完全知道马尔可夫链的统计特性.

上面我们介绍了马尔可夫链的定义. 对一般的随机过程要定义它的马尔可夫性, 需要一般的条件概率、条件分布的概念, 这些已超出本书范围, 因此以下的叙述是直观的、不太严格的.

随机过程$\{X(t), t\in T\}$, 如果对任意的$t_1<t_2<\cdots<t_n$及$t_i\in T$, $X(t_n)$关于$X(t_1)$, $X(t_2)$, $\cdots$, $X(t_{n-1})$的条件分布恰好等于$X(t_n)$关于$X(t_{n-1})$的条件分布(关于条件分布可参见第一册§3.2的解释), 即$P\{X(t_n)<y|X(t_{n-1})=x_{n-1},\cdots,X(t_1)=x_1\}=P\{X(t_n)<y|X(t_{n-1})=x_{n-1}\}$则称$\{X(t), t\in T\}$具有**马尔可夫性**或**称它为马尔可夫过程**. 上式是在一般情况下表明当$X(t)$"现在"已知时, "过去"和"将来"的统计特性是无关的.

马尔可夫过程$\{X(t), t\in T\}$所可能取的值的全体, 称为过程的**状态空间**, 其中每一元素, 即$\{X(t), t\in T\}$所可能取的值称为**状态**. $X(t)=x$, 常说为: 在时刻t时, 过程(或系统)在状态x中. 本书中所研究的马尔可夫过程, 其状态空间或是$(-\infty, \infty)$, 或是它的某个子集, 如$\{0, 1, 2, \cdots\}$, $[0, \infty)$等.

对于马尔可夫过程$\{X(t), t\in T\}$来说, 最重要的概念是**转移概**

率 $P(s,x;t,A)$，其中 $s,t\in T$，且 $s<t$，A 是状态空间的子集，它表示在 $X(s)=x$ 的条件下，$X(t)\in A$ 的条件概率，即

$$P(s,x;t,A)=P\{X(t)\in A|X(s)=x\}$$

特别当 $A=(-\infty,y)$ 时，有

$$\begin{aligned}&P(s,x;t,(-\infty,y))\\&=P\{X(t)\in(-\infty,y)|X(s)=x\}\\&=P\{X(t)<y|X(s)=x\}\end{aligned}$$

这时为方便起见，也常把 $P(s,x;t,(-\infty,y))$ 简记为 $F(s,x;t,y)$，而且今后更常用的是 $F(s,x;t,y)$，称它为转移概率分布。对转移概率分布 $F(s,x;t,y)$ 也有如下的类似于(4)的切普曼-柯尔莫哥洛夫方程

$$F(s,x;t,y)=\int_{-\infty}^{\infty}d_zF(s,x;u,z)\cdot F(u,z;t,y),\qquad s<u<t$$

上式的证明也可象(4)的证明一样，按 $X(u)$ 所处的不同状态运用全概率公式，并利用 $X(t)$ 的马尔可夫性，把有关的条件概率用转移概率写出即可。

对于一般马尔可夫过程，也象马尔可夫链一样，它所有的有限维分布都可用它在初始时刻 $t=a$ 时 $X(a)$ 的分布和转移概率分布来表示，因此，只要知道了马尔可夫过程的初始分布和转移概率，那么也就知道了它的有限维分布，亦即完全知道了马尔可夫过程的统计特性。

转移概率分布 $F(s,x;t,y)$ 按定义关于 y 是一个分布函数，因而有：

(i) $F(s,x;t,y)\geqslant 0$；

(ii) $F(s,x;t,\infty)=1$；

(iii) $F(s,x;t,y)$ 关于 y 是单调不减和左连续的。

而 $F(s,x;t,y)$ 对固定的 y,s,t 是 x 的波雷尔可测函数。

如果 $F(s,x;t,y)$ 关于 y 的导数存在，则有

$$p(s,x;t,y)=\frac{\partial}{\partial y}F(s,x;t,y)$$

称为马尔可夫过程的**转移密度**.

在一些重要的特殊场合，$F(s,x;t,y)$ 只依赖于 $t-s$，而不依赖于 s,t 的本身，这时一般记

$$F(t-s,x;y)=F(s,x;t,y)$$

称为**齐次的转移概率**，相应的马尔可夫过程也称为**齐次的**. 按上述定义，可知 $F(\tau,x;y)$ 表示在某时刻 t，过程在状态 x，而经过 τ 这段时间之后，过程位于 $(-\infty,y)$ 中的概率，相应的转移密度记为 $p(\tau,x;y)$.

所以马尔可夫链就是参数 t 只取离散值的马尔可夫过程，此时参数集 T 一般取为 $\{0,1,2,\cdots\}$.

如果马尔可夫链的状态空间也是离散的，（例如状态空间是 $\{0,\pm1,\pm2,\cdots\}$）就称它为**可数（状态的）马尔可夫链**. 如果状态空间还是有限的，就称为**有限（状态的）马尔可夫链**. 这是最简单的马尔可夫过程.

在可数马尔可夫链的场合，还常引入所谓的**转移矩阵**.

$$P^{(n)}(m)\triangleq\{p_{ij}^{(n)}(m)\}_{i,j\in I},\ m\in T$$

于是切普曼-柯尔莫哥洛夫方程可简洁地表示为：

$$P^{(l+n)}(m)=P^{(l)}(m)\cdot P^{(n)}(m+l)$$

如再记

$$P(m)=P^{(1)}(m)=\{p_{ij}(m)\}_{i,j\in I}$$

则有

$$P^{(n)}(m)=P(m)P(m+1)\cdots P(m+n-1)$$

在齐次的场合，有

$$P^{(n)}(1)=P^{(n)}(2)=\cdots\triangleq P^{(n)}=\{p_{ij}^{(n)}\}_{i,j\in I}$$

相应的切普曼-柯尔莫哥洛夫方程为:

$$p_{ij}^{(l+n)}=\sum_{k\in I}p_{ik}^{(l)}p_{kj}^{(n)}$$

或者

$$P^{(l+n)}=P^{(l)}\cdot P^{(n)}$$

同样，如记

$$P=P^{(1)}=\{p_{ij}\}_{i,j\in I}$$

则有

$$\begin{aligned}P^{(n)}&=P^{(n-1+1)}=P^{(n-1)}P\\&=P^{(n-2)}\cdot P\cdot P=P^{(n-2)}\cdot P^2\\&=\cdots=P^n\end{aligned}$$

其中 P^n 表示矩阵 P 的 n 次方.

为了帮助理解上述概念,我们下面给出一些简单的例子,更多的内容将在第二章中叙述.

[例 1]　**无限制随机游动**　考虑在直线上作随机游动的质点,如果某个时刻质点位于 i，则在下一步质点以概率 p 向右移动一格到 $i+1$,而以概率 $q=1-p$ 向左移动一格到 $i-1$. 若以 $X(n)$ 表时刻 n 质点的位置,那么 $\{X(n),\ n=0,\ 1,\ 2,\cdots\}$ 是一个随机过程,而且当 $X(n)=i$ 时, $X(n+1)$, $X(n+2)$, $\cdots$, $X(n+k)$ 等以后的行为只与 $X(n)=i$ 有关,而与质点在 n 以前是如何到 i 是完全无关的. 所以它是一个马尔可夫链，而且还是齐次的. 其状态空间 $I=\{\cdots-2,-1,0,1,2,\cdots\}$,一步转移概率为:

$$p_{i\,i+1}=p,\quad p_{i\,i-1}=q=1-p,\quad i\in I$$

其中 $0\leqslant p\leqslant 1$,其它的 p_{ij} 都等于 0.

我们来求这个马尔可夫链的高阶转移概率 $p_{ij}^{(n)}$. 由于共进行了 n 次转移，每次转移只有两种可能的选择：向左，其概率为 q;

向右，其概率为 p. 而 n 次转移的结果是从 i 到了 j. 因此 n 次转移中向右的次数减去向左的次数应等于 $j-i$. 于是可知：

$$p_{ij}^{(n)}=\begin{cases}\begin{pmatrix} n \\ \frac{n+j-i}{2}\end{pmatrix} p^{(n+j-i)/2}\cdot q^{(n-j+i)/2}, & \text{当 } n+j-i \text{ 为偶数}\\ 0, & \text{当 } n+j-i \text{ 为奇数}\end{cases}$$

特别有

$$p_{ii}^{(n)}=\begin{cases}\begin{pmatrix} n \\ \frac{n}{2}\end{pmatrix}\cdot (pq)^{\frac{n}{2}}, & n \text{ 是偶数}\\ 0, & n \text{ 是奇数}\end{cases}$$

利用斯特灵(Stirling)公式 $n!\doteq\left(\frac{n}{e}\right)^n\cdot\sqrt{2\pi n}$, 可得

$$p_{ii}^{(2n)}=\begin{pmatrix} 2n \\ n\end{pmatrix}(pq)^n=\frac{(2n)!}{(n!)(n!)}(pq)^n$$

$$\doteq\frac{\left(\frac{2n}{e}\right)^{2n}\cdot\sqrt{2\pi\cdot 2n}}{\left(\frac{n}{e}\right)^n\sqrt{2\pi n}\cdot\left(\frac{n}{e}\right)^n\sqrt{2\pi n}}(pq)^n$$

$$=\frac{(4pq)^n}{\sqrt{\pi n}}$$

［例 2］ **带有一个吸收壁的随机游动** 仍考虑随机游动，但其状态空间 $I=\{0,1,2,\cdots\}$，而且一旦当 $X(n)=0$ 以后，$X(n+1)$ 也就停留在 0 这个状态上，这样的状态称为**吸收状态**. 同例 1 一样的道理，$\{X(n), n=0,1,\cdots\}$ 也是一个齐次马尔可夫链. 它的一步转移概率为：

$$p_{i\,i+1}=p,\quad p_{i\,i-1}=q=1-p,\quad i\geqslant 1,\quad p_{00}=1$$

在本例中，0 正好又是状态空间的端点(壁)，所以就称这一随机游动为带有一个吸收壁的随机游动. 不难看出，i 是马尔可夫链吸收

状态的充要条件是 $p_{ii}=1$.

[例 3] **带有两个吸收壁的随机游动** 若对随机游动取状态空间 $I=\{0,1,2,\cdots,a\}$, 且 0, a 为它的吸收状态, 则它也是齐次马尔可夫链, 而一步转移概率为:

$$p_{i\,i+1}=p,\ \ p_{i\,i-1}=q=1-p,\ \ 1\leqslant i\leqslant a-1$$

$$p_{00}=p_{aa}=1$$

[例 4] **带有一个反射壁的随机游动** 对随机游动取状态空间 $I=\{0,1,2,\cdots\}$, 这里和例 2 不同, 0 这个状态不再是吸收状态了, 但粒子一旦进入状态 0 后, 下一步它以概率 p 向右移一格, 以概率 q 停在 0 状态. 这后一种情况也可解释为: 设想在"$-\frac{1}{2}$"处有一反射壁, 每次粒子自 0 向左移动时, 即被反射回状态 0. 正是基于上述解释, 我们把这个随机游动称为带有一个反射壁的随机游动. 它仍是一个齐次马尔可夫链, 其一步转移概率为:

$$p_{i\,i+1}=p,\ \ p_{i\,i-1}=q=1-p,\ \ i\geqslant 1$$

$$p_{00}=q,\ \ p_{01}=p$$

[例 5] **带有两个反射壁的随机游动** 此时状态空间 $I=\{0,1,2,\cdots,a\}$, 一步转移概率为:

$$p_{i\,i+1}=p,\ \ p_{i\,i-1}=q=1-p,\ \ 1\leqslant i\leqslant a-1$$

$$p_{00}=q,\ p_{01}=p;\ p_{aa}=p,\ p_{a\,a-1}=q$$

仿前可设想在"$-\frac{1}{2}$"和"$a+\frac{1}{2}$"处分别有一个反射壁.

[例 6] **爱伦菲斯特(Ehrenfest) 模型** 这是一个齐次的有限马尔可夫链, 其状态空间为 $I=\{-a,\ -a+1,\ \cdots,\ -1,0,1,2,\cdots,a\}$, 一步转移概率是:

$$p_{i\,i-1}=\frac{1}{2}\left(1+\frac{i}{a}\right),\ p_{i\,i+1}=\frac{1}{2}\left(1-\frac{i}{a}\right),\qquad -a\leqslant i\leqslant a$$

$$p_{a\,a-1}=1,\ p_{-a\,-a+1}=1$$



由定义可以看出，当系统所处的状态 i 在原点的左边时，即 $i<0$ 时，系统在下一步向左移的概率

$$p_{i\,i-1}=\frac{1}{2}\left(1+\frac{i}{a}\right)<\frac{1}{2}$$

而向右移的概率

$$p_{i\,i+1}=\frac{1}{2}\left(1-\frac{i}{a}\right)>\frac{1}{2}$$

亦即当系统在原点左边时，系统下一步更倾向于（即以更大的概率）向右移，且其倾向性与离原点的距离成正比．类似地，当系统在原点右边时，系统下一步将更倾向于向左移．特别当系统在原点时，系统以等概率$\left(\frac{1}{2}\right)$向左或向右移；而在 $-a$ 处则以概率 1 向右移；在 a 处则以概率 1 向左移．

这样的一个随机游动可以有如下两种物理的解释．

第一种解释　考虑一容器，其中有 $2a$ 个粒子在作随机的杂乱运动．设想有一个并不存在的界面，将这容器分为相等的左、右两部份．均衡时，每一部份中都包含 a 个粒子．现在如用 $Y(t)$ 表示时刻 t 时，右边那部份中所含粒子数与 a 的差，则 $Y(t)$ 显然是随机变量，且只能取值：$-a, -a+1, \cdots, -2, -1, 0, 1, 2, \cdots, a-1, a$．如有一个粒子从左边跑到右边，则 $Y(t)$ 增加 1；反之，如有一粒子从右边跑到左边，则 $Y(t)$ 减少 1．作为第一步近似，如假定，同时两个粒子从左边跑到右边，或从右边跑到左边，或一个从右边到左边，一个从左边到右边等都不可能发生，亦即 $Y(t)$ 只能发生增加 1 或减少 1 的变化．此时如用 $X(n)$ 表示发生 n 次变化后的两边粒子数的差，则它恰好可用上述的爱伦菲斯特模型来表示：当 $i>0$，即右边粒子比左边更多时，由于粒子的运动是随机的，因而从右边向左边跑去一个粒子的概率要比从左边向右边跑去一个粒子的概率大，而且大的程度应与双边粒子数的差：$(i+a)-[2a-$

$(i+a)]=2i$ 成正比，即与 i 成正比，当 $i<0$ 时，可同样解释.

第二种解释 设想一粒子在直线上作随机游动，但同时此质点又与原点用一"弹簧"相联，因而当质点偏离原点时，即受一附加的与偏离成正比的向原点的力的作用，从而导致质点向原点移动的概率增大. 用 $X(n)$ 表示时间 n 时粒子的位置，则可用上述爱伦菲斯特模型来近似地描述它.

[例 7] （非齐次马尔可夫链） 考虑卜里耶(Polya)模型(参见第一册例 2.1.2)，罐中含有 b 只黑球及 r 只红球，每次随机地取出一只后把原球放回，并加入与抽出球同色的球 c 只，再摸第二次并继续进行下去. 以罐中黑球数为状态，而"$X(n)=i$"表示在第n 次摸球放回后在罐中有 i 个黑球这一事件. 显然每摸一次后罐中黑球数或者不变，或者增加 c 个. 所以

$$P\{X(n+1)=j|X(n)=i,A\}=\begin{cases}\dfrac{i}{b+r+nc}, & \text{若 } j=i+c\\ 1-\dfrac{i}{b+r+nc}, & \text{若 } j=i\\ 0, & \text{其它}\end{cases}$$

其中 A 表示由前 $n-1$ 次摸球结果规定的任一事件. 由于上述概率只与 i, j 和 n 有关，而与 A 无关，所以它是一个马尔可夫链，其转移概率为

$$p_{ij}(n)=\begin{cases}\dfrac{i}{b+r+nc}, & j=i+c\\ 1-\dfrac{i}{b+r+nc}, & j=i\end{cases}$$

因此它还是非齐次的.

[例 8] **离散分支过程** 考虑某一群体. 用 $X(n)$ 表示第 n 代中个体的数目. 假定某一代的每一个个体可产生 Y 个下一代个体，其中 Y 是一个取非负整数值的离散随机变量:

$$P\{Y=k\}=p_k,\quad k=0,1,2,\cdots$$

并设某一代中各个体产生下一代的个数相互独立，则有

$$X(n+1)=Y_1+Y_2+\cdots+Y_{X(n)}$$

其中$\{Y_i\}$是相互独立与Y同分布的随机变量. 由此可知，只要知道了$X(n)$，那末$X(n+1)$的分布就完全决定了，而与以前(即更前的代的个体数$X(n-1)$，$X(n-2)$，…)无关，从而是一个马尔可夫链. 这种马尔可夫链，称为**离散的分支过程**。

如引入母函数

$$G(Z)\triangleq\sum_{k=0}^{\infty}p_kZ^k$$

$$F_n(Z)\triangleq\sum_{k=0}^{\infty}P\{X(n)=k\,|\,X(0)=1\}\cdot Z^k$$

$$=\sum_{k=0}^{\infty}p_{1k}^{(n)}\cdot Z^k$$

则根据母函数的性质(参见第一册(4.3.19))可得:

$$F_{n+1}(Z)=F_n(G(Z))=F_{n-1}(G(G(Z)))=\cdots$$

显然有$F_0(Z)=Z$，所以

$$F_n(Z)=F_0(G^{(n)}(Z))=G^{(n)}(Z)$$

其中$G^{(n)}(Z)$表示$\underbrace{G(G(\cdots G(Z)\cdots))}_{n\text{次}}$.

一般有

$$\sum_{k=0}^{\infty}p_{ik}^{(n)}Z^k=\left(\sum_{k=0}^{\infty}p_{1k}^{(n)}\cdot Z^k\right)^i$$

$$=(F_n(Z))^i=(G^{(n)}(Z))^i$$

由此可知，$G(Z)$，即$\{p_k, k=0,1,2,\cdots\}$，完全决定了$\{X(n)\}$的分布.

如果假定$\{p_k\}$的均值存在:

$$m=\sum_{k=1}^{\infty}k\cdot p_k<\infty$$

则它表示每个个体产生下一代个体的平均个数，直观上很清楚，如果$m<1$，即每个个体产生下一代的平均个数不到一个，则整个群体将灭种，即$X(n)$迟早将变为0，从而一直为0（即0是吸收状态）；如果$m>1$，则整个群体将不断增长，但也有一定的概率灭种，如果$m=1$，直观上是不很清楚的，理论上可以证明，此时群体迟早将灭种，关于这些结果的证明将在第二章叙述。

一种与马尔可夫过程密切有关的是独立增量过程。一个随机过程$\{X(t),\ t\in T\}$称为**独立增量**的，如果对任意的$t_1<t_2<\cdots<t_{2n-1}<t_{2n}$，过程的增量$X(t_2)-X(t_1)$，$\cdots$，$X(t_{2n})-X(t_{2n-1})$是相互独立的随机变量。若参数集$T=[a,\ b)$，则象马尔可夫过程一样，独立增量过程的有限维分布可由它的初始分布$P\{X(a)<x\}$和一切增量的分布唯一确定。

对于独立增量过程$\{X(t),t\in T\}$，当$T=[a,b)$，$a>-\infty$时，一般总规定$P\{X(a)=0\}=1$。这时对任意的$t_1<t_2<\cdots<t_n<t$，$X(t)-X(t_n)$与$X(t_1)$ $(=X(t_1)-X(a))$，$X(t_2)-X(t_1)$，$\cdots$，$X(t_{n-1})-X(t_{n-2})$独立，所以当已知$X(t_n)$时，$X(t)(=X(t)-X(t_n)+X(t_n))$与$X(t_1)$，$X(t_2)(=X(t_2)-X(t_1)+X(t_1))$，$\cdots$，$X(t_{n-1})$独立，因此$\{X(t),t\in T\}$一定是一个马尔可夫过程。

如果独立增量过程$\{X(t),\ t\in T\}$还满足$E\{X(t)\}=0$，$E\{|X(t)|^2\}<\infty$，则它也是正交增量过程。事实上，当$t_1<t_2<t_3<t_4$时我们有

$$\begin{aligned}&E\{(X(t_2)-X(t_1))\overline{(X(t_4)-X(t_3))}\}\\&=E\{X(t_2)-X(t_1)\}\overline{E\{X(t_4)-X(t_3)\}}\\&=0\end{aligned}$$

亦即均值函数为零的二阶矩过程，如同时又是独立增量过程，则它

也是正交增量过程.

如果增量$X(t_2)-X(t_1)$的分布只与t_2-t_1有关，而与t_2, t_1本身无关，则称为**齐次的**.

两个最重要的独立增量过程是维纳(Wiener)-爱因斯坦(Einstein)过程和普阿松(Poisson)过程.

维纳-爱因斯坦过程 它也是实正态过程，其均值函数为零，而其协方差函数为

$$\Gamma(s,t)=E\{X(s)X(t)\}=\sigma^2\cdot\min(s,t)$$

由此可求出其增量$X(t)-X(s)$的方差(假定$s<t$)：

$$\begin{aligned}D(X(t)-X(s))&=E|X(t)-X(s)|^2\\&=E\{(X(t)-X(s))(X(t)-X(s))\}\\&=E\{X(t)X(t)\}-E\{X(t)X(s)\}\\&\quad-E\{X(s)X(t)\}+E\{X(s)X(s)\}\\&=\Gamma(t,t)-\Gamma(t,s)-\Gamma(s,t)+\Gamma(s,s)\\&=\sigma^2\cdot t-\sigma^2\cdot s-\sigma^2\cdot s+\sigma^2\cdot s=\sigma^2\cdot(t-s)\end{aligned}$$

一般地有：

$$D(X(t)-X(s))=\sigma^2\cdot|t-s|$$

普阿松过程 这是一个实的独立增量过程，$X(0)=0$，其增量$X(t)-X(s)$ $(s<t)$的分布是普阿松分布：

$$P\{X(t)-X(s)=j\}=e^{-\lambda(t-s)}\cdot\frac{(\lambda(t-s))^j}{j!},\quad j=0,1,2,\cdots$$

不难算出它增量的均值和方差为

$$\begin{aligned}E\{X(t)-X(s)\}&=D\{X(t)-X(s)\}\\&=\lambda(t-s)\end{aligned}$$

而过程本身的均值函数为

$$m(t)=E\{X(t)\}=E\{X(t)-X(0)\}=\lambda t$$

当$s<t$时其协方差函数为

$$\begin{aligned}\Gamma(s,t)&=E\{(X(s)-\lambda s)(X(t)-\lambda t)\}\\&=E\{(X(s)-\lambda s)(X(s)-\lambda s)\\&\quad+[(X(t)-\lambda t)-(X(s)-\lambda s)]\}\\&=E\{(X(s)-\lambda s)^2\}-E\{X(s)-\lambda s\}\\&\quad\cdot E\{(X(t)-\lambda t)-(X(s)-\lambda s)\}\\&=\lambda s\end{aligned}$$

一般地，过程的协方差函数可表为

$$\Gamma(s,t)=\lambda\cdot\min(s,t)$$

由此可见，这两个过程都是齐次的，但都不是平稳的.

三、其它 除上面介绍的二阶矩过程（包括其特殊情形平稳过程，正态过程，正交增量过程等）和马尔可夫过程（包括其特殊情形马尔可夫链，分支过程，独立增量过程等）外，还有一些比较重要的随机过程，例如“鞅”，点过程等.

此外，我们现在讨论的随机过程$\{X(t),\ t\in T\}$，其中每个$X(t)$都是随机变量，如果它是随机向量，那就是向量随机过程；而且参数集T也限于直线的某个子集，如果T可以是平面上的某个子集，相应地每个t是平面上的点（要有两个坐标），那就是**随机场**. T当然还可以是三维的，四维的等等，但仍称为随机场. 关于这两方面，在本书中也都不讨论了.

第二章　马尔可夫过程

§1. 马尔可夫链

在本节中我们限于研究齐次的可数马尔可夫链，为确定起见，设所研究的齐次可数马尔可夫链为$\{X(n), n=0, 1, 2, \cdots\}$，并且以$I=\{0, 1, 2, \cdots\}$为它的状态空间．今后，还将把"$X(n)=i$"说为"时刻$n$时马尔可夫链或系统处于状态$i$"等等．

再设此马尔可夫链的一步转移概率为$\{p_{ij}\}$；n步转移概率为$\{p_{ij}^{(n)}\}$．

这一节分为四小段．第一、二段从研究状态转移的情况出发，依次引出系统状态的分类，以及状态空间的分解等等内容．第三段研究状态运动的极限性质．第四段进一步讨论一些特殊的马尔可夫链，并指出某些应用．

一、首次进入时间和状态分类

如果对状态i和j，存在某个$n\geqslant 1$，使$p_{ij}^{(n)}>0$，即从状态i出发，经某个n步，可以正的概率到达状态j，则称**自状态i可达状态j**，并记为$i\to j$．反之，如自状态i不可达状态j，记为$i\nrightarrow j$，此时对一切$n\geqslant 1$，$p_{ij}^{(n)}=0$．

如$i\to j$，$j\to i$，则称**状态i和状态j相通**，记为$i\longleftrightarrow j$．我们有下列的

定理1　如$i\to k$，$k\to j$，则$i\to j$．

[证明]　如果$i\to k$，$k\to j$，则按定义存在$l\geqslant 1$和$n\geqslant 1$使$p_{ik}^{(l)}>0$，$p_{kj}^{(n)}>0$，根据切普曼-柯尔莫哥洛夫方程，有

$$p_{ij}^{(l+n)}=\sum_{m=0}^{\infty}p_{im}^{(l)}p_{mj}^{(n)}\geqslant p_{ik}^{(l)}p_{kj}^{(n)}>0,\quad k\in I \tag{1}$$

其中第一个不等式是由于所有的$p_{ik}^{(l)}$和$p_{kj}^{(n)}$都是非负的，按定义，即 $i\to j$.

系 1 如 $i\to k_1, k_1\to k_2, \cdots, k_{n-1}\to k_n, k_n\to j$, 则 $i\to j$.

系 2 如 $i\longleftrightarrow k_1, k_1\longleftrightarrow k_2, \cdots, k_{n-1}\longleftrightarrow k_n, k_n\longleftrightarrow j$, 则 $i\longleftrightarrow j$.

应用上述定理和系，不难看出，第一章中的那些随机游动中，所有的状态只要不是吸收状态，它与自己相邻的非吸收状态都是相通的，从而在不带吸收壁的随机游动中，所有的状态都是相通的，而在带有吸收壁的随机游动中，除吸收状态外，余下的状态也都是相通的. 收吸状态不与任何别的状态相通，而且也不能自吸收状态到达任何别的状态，但在带吸收壁的随机游动的场合，从任何非吸收状态都可到达吸收状态.

对于任意两个状态 i 和 j，在事件 $\{X(0)=i\}$ 上引入随机变量

$$T_{ij}\triangleq\min\{n: X(0)=i, X(n)=j, n\geqslant 1\}$$

即 T_{ij} 是从状态 i 出发，首次进入状态 j 的时刻，对$\omega\in\{X(0)\neq i\}$或 $\omega\in\{X(0)=i,\ X(n)\neq j,\ n\geqslant 1\}$，我们也可形式地补充规定为 $T_{ij}(\omega)=+\infty$. 此外，再定义

$$f_{ij}^{(n)}\triangleq P\{T_{ij}=n\,|\,X(0)=i\}\geqslant 0$$

即系统自状态 i 出发，经 n 步首次到达状态 j 的概率，显然有：

$$f_{ij}^{(n)}=P\{X(n)=j; X(m)\neq j, m=1, 2, \cdots, n-1\,|\,X(0)=i\}$$

$$=\sum_{i_1\neq j}\cdots\sum_{i_{n-1}\neq j}p_{ii_1}p_{i_1i_2}\cdots p_{i_{n-1}j}\qquad (n\geqslant 1)$$

再定义

$$f_{ij}\triangleq\sum_{n=1}^{\infty}f_{ij}^{(n)}=\sum_{n=1}^{\infty}P\{T_{ij}=n\}=P\{T_{ij}<\infty\}$$

它是系统自状态 i 出发，迟早要到达状态 j 的概率，显然有

$$0 \leqslant f_{ij}^{(n)} \leqslant f_{ij} \leqslant 1$$

我们先来证明一个联系 $f_{ij}^{(n)}$ 和 $p_{ij}^{(n)}$ 的常用引理，

引理 对任意的 $i, j \in I$ 及 $1 \leqslant n < \infty$，有

$$p_{ij}^{(n)} = \sum_{l=1}^{n} f_{ij}^{(l)} p_{jj}^{(n-l)} \tag{2}$$

[证明] 首先由于 $\{X(0)=i, X(n)=j\} \subset \{T_{ij} \leqslant n, X(n)=j\}$，因此对事件 $\{X(n)=j\}$ 可按照 $X(n)$ 首次进入 j 的时刻不同而分解为互不相容事件之和，即

$$\begin{aligned}
p_{ij}^{(n)} &= P\{X(n)=j \mid X(0)=i\} \\
&= P\{T_{ij} \leqslant n, X(n)=j \mid X(0)=i\} \\
&= \sum_{l=1}^{n} P\{T_{ij}=l, X(n)=j \mid X(0)=i\} \\
&= \sum_{l=1}^{n} P\{T_{ij}=l \mid X(0)=i\} P\{X(n)=j \mid T_{ij} \\
&= l, X(0)=i\} \\
&= \sum_{l=1}^{n} P\{T_{ij}=l \mid X(0)=i\} P\{X(n) \\
&= j \mid X(0)=i, X(1) \neq j, \cdots, X(l-1) \neq j, X(l)=j\} \\
&= \sum_{l=1}^{n} P\{T_{ij}=l \mid X(0)=i\} P\{X(n)=j \mid X(l)=j\} \\
&= \sum_{l=1}^{n} f_{ij}^{(l)} p_{jj}^{(n-l)}
\end{aligned}$$

在上述推导中，第五个等号是用了 T_{ij} 的定义，而第六个等号是用了 $X(n)$ 的马尔可夫性，因为 $\{X(0)=i, X(1) \neq j, \cdots, X(l-1) \neq j\}$

为一 l 以前发生的事件，在 $X(l)=j$ 的条件下，它与 $X(n)=j$ 独立，最后一个等式是用到了 $f_{ij}^{(l)}$ 的定义和 $X(n)$ 的齐次性．于是引理得证．

利用上述引理，可以帮助我们建立 $p_{ij}^{(n)}$ 的许多性质．

定理 2　$f_{ij}>0$ 的充要条件是 $i\to j$.

［证明］　**充分性**　如 $i\to j$，则有某 $n\geqslant 1$ 使 $p_{ij}^{(n)}>0$，由(2)

$$\sum_{l=1}^{n} f_{ij}^{(l)} p_{jj}^{(n-l)} = p_{ij}^{(n)} > 0$$

从而 $f_{ij}^{(1)}, f_{ij}^{(2)}, \cdots, f_{ij}^{(n)}$ 中至少有一个为正的，这样

$$f_{ij} = \sum_{l=1}^{\infty} f_{ij}^{(l)} > 0$$

必要性　如果 $f_{ij}>0$，则因 $f_{ij}=\sum_{n=1}^{\infty} f_{ij}^{(n)}$，所以至少有一个 $n\geqslant 1$，使 $f_{ij}^{(n)}>0$．另外，由(2)得

$$p_{ij}^{(n)} = \sum_{l=1}^{n} f_{ij}^{(l)} p_{jj}^{(n-l)} \geqslant f_{ij}^{(n)} p_{jj}^{(0)}$$
$$= f_{ij}^{(n)} > 0$$

于是 $i\to j$.

系　$i\longleftrightarrow j$，即 i 与 j 相通的充要条件是：$f_{ij}>0, f_{ji}>0$.

在 $i=j$ 时，T_{ii} 便是自状态 i 出发，首次返回状态 i 的时刻，f_{ii} 便是自状态 i 出发，在有限步内迟早要返回状态 i 的概率．它当然是 0 与 1 之间的一个数，但可能有两种情况．

如果 $f_{ii}=1$，则称状态 i 是**常返的**；如果 $f_{ii}<1$，则称状态 i 是非常返的，有时也称为**滑过的**．

下面的定理，说明上述定义的合理性．

定理 3　如状态 j 是常返的，则以概率 1，系统无穷次返回状态 j；如状态 j 是非常返的，则以概率1，系统只有有限次返回状态

j，亦即系统无穷次返回状态 j 的概率为零.

[证明] 用 $Q_{ij}(m)$ 表示自状态 i 出发，系统至少进入状态 j m 次的概率；Q_{ij} 表示自状态 i 出发，系统无穷多次进入状态 j 的概率，则有

$$\lim_{m\to\infty} Q_{ij}(m) = Q_{ij}$$

另一方面，我们有

$$Q_{ij}(1) = f_{ij}$$

仿定理 2 证明中的方法，利用过程的马尔可夫性，有

$$\begin{aligned} Q_{ij}(m+1) &= \sum_{n=1}^{\infty} f_{ij}^{(n)} \cdot Q_{jj}(m) \\ &= f_{ij} \cdot Q_{jj}(m) \end{aligned}$$

反复应用上式，得

$$\begin{aligned} Q_{ij}(m+1) &= f_{ij} \cdot Q_{jj}(m) = f_{ij} \cdot f_{jj} \cdot Q_{jj}(m-1) \\ &= \cdots = f_{ij} \cdot (f_{jj})^m \end{aligned}$$

于是有：

$$Q_{jj} = \lim_{m\to\infty} Q_{jj}(m) = \begin{cases} 1, & \text{如 } f_{jj}=1, \text{即 } j \text{ 常返} \\ 0, & \text{如 } f_{jj}<1, \text{即 } j \text{ 非常返} \end{cases}$$

系 $$Q_{ij} = \begin{cases} f_{ij}, & j \text{ 常返} \\ 0, & j \text{ 非常返} \end{cases}$$

下面的定理给出基于 $p_{jj}^{(n)}$ 的状态常返与否的判别准则.

定理 4 状态 j 常返的充要条件是 $\sum_{n=1}^{\infty} p_{jj}^{(n)} = \infty$.

[证明] 由(2)，特别取 $i=j$，有

$$p_{jj}^{(n)} = \sum_{l=1}^{n} f_{jj}^{(l)} p_{jj}^{(n-l)}$$

两边对 n 相加，得

$$\sum_{n=1}^{N} p_{jj}^{(n)} = \sum_{n=1}^{N} \sum_{l=1}^{n} f_{jj}^{(l)} p_{jj}^{(n-l)}$$

$$=\sum_{l=1}^{N}\sum_{n=l}^{N}f_{jj}^{(l)}p_{jj}^{(n-l)}$$

$$=\sum_{l=1}^{N}f_{jj}^{(l)}\cdot\sum_{m=0}^{N-l}p_{jj}^{(m)} \tag{3}$$

$$\leqslant\sum_{l=1}^{N}f_{jj}^{(l)}\cdot\sum_{m=0}^{N}p_{jj}^{(m)}$$

于是

$$\frac{\sum_{n=1}^{N}p_{jj}^{(n)}}{\sum_{m=0}^{N}p_{jj}^{(m)}}=\frac{\sum_{n=1}^{N}p_{jj}^{(n)}}{1+\sum_{m=1}^{N}p_{jj}^{(m)}}\leqslant\sum_{l=1}^{N}f_{jj}^{(l)}$$

令 $N\to\infty$，得：

$$\lim_{N\to\infty}\frac{\sum_{n=1}^{N}p_{jj}^{(n)}}{1+\sum_{n=1}^{N}p_{jj}^{(n)}}\leqslant f_{jj}\leqslant 1$$

如$\sum_{n=1}^{\infty}p_{jj}^{(n)}=\infty$，则易知上式左边极限为1，于是 $f_{jj}=1$，即 j 常返，充分性证毕。

现证必要性，在(3)中，如取 $N'<N$，则有

$$\sum_{n=1}^{N}p_{jj}^{(n)}\geqslant\sum_{l=1}^{N'}f_{jj}^{(l)}\sum_{m=0}^{N-l}p_{jj}^{(m)}$$

$$\geqslant\sum_{l=1}^{N'}f_{jj}^{(l)}\sum_{m=0}^{N-N'}p_{jj}^{(m)}$$

于是

$$\frac{\sum_{n=1}^{N}p_{jj}^{(n)}}{\sum_{m=0}^{N-N'}p_{jj}^{(m)}}=\frac{\sum_{n=1}^{N}p_{jj}^{(n)}}{1+\sum_{m=1}^{N-N'}p_{jj}^{(m)}}\geqslant\sum_{l=1}^{N'}f_{jj}^{(l)}$$

如设$\sum_{n=1}^{\infty} p_{jj}^{(n)} < \infty$，则在上式中令$N \to \infty$，得

$$1 > \frac{\sum_{n=1}^{\infty} p_{jj}^{(n)}}{1+\sum_{n=1}^{\infty} p_{jj}^{(n)}} \geqslant \sum_{l=1}^{N'} f_{jj}^{(l)}$$

再令$N' \to \infty$，得：

$$1 > \frac{\sum_{n=1}^{\infty} p_{jj}^{(n)}}{1+\sum_{n=1}^{\infty} p_{jj}^{(n)}} \geqslant f_{jj}$$

即j是非常返状态．这与必要性证明的假定矛盾，所以$\sum_{n=1}^{\infty} p_{jj}^{(n)}$不能小于$\infty$，即$\sum_{n=1}^{\infty} p_{jj}^{(n)} = \infty$。

对于常返状态i，我们进一步来研究它的**平均返回时间**：

$$\mu_i = \sum_{n=1}^{\infty} n \cdot f_{ii}^{(n)}$$

同样可能有两种情形：如果$\mu_i < \infty$，则称i**为正常返的**；如果$\mu_i = \infty \left(\text{或} \frac{1}{\mu_i} = 0\right)$，则称$i$为**零常返的**．

称状态i是**周期为$t(t>1)$的**，如果$p_{ii}^{(n)}$除$n=t, 2t, 3t, \cdots$外均为零，但没有比t更大的t'使$p_{ii}^{(n)}$除$n=t', 2t', 3t', \cdots$外均为零；当不存在上述的t时，状态i称为**非周期的**．

非周期的正常返状态，称为**遍历状态**．

我们不加证明地给出下列两条定理．

定理 5　如i是常返状态，则i是零常返的充要条件是：

$$\lim_{n \to \infty} p_{ii}^{(n)} = 0$$

系 如果 j 是零常返状态，则 $\lim\limits_{n\to\infty} p_{ij}^{(n)}=0$.

[证明] 由(2)式，得

$$p_{ij}^{(n)}=\sum_{l=1}^{n} f_{ij}^{(l)} p_{jj}^{(n-l)}$$

$$\leqslant\sum_{l=1}^{n'} f_{ij}^{(l)} p_{jj}^{(n-l)}+\sum_{l=n'+1}^{n} f_{ij}^{(l)}$$

由于 j 是零常返状态，$p_{jj}^{(n)}\to 0$，于是固定 n' 后，令 $n\to\infty$ 时，右边第一项将趋于零. 同时，由于 $\sum\limits_{l=1}^{\infty} f_{ij}^{(l)}=f_{ij}\leqslant 1$ 是收敛的正项级数，所以其余项之和 $\sum\limits_{l=n'+1}^{\infty} f_{ij}^{(l)}$ 在 $n'\to\infty$ 时应趋于零. 系由此得证.

定理 6[①] 如 i 是遍历状态，即非周期的正常返状态，则

$$\lim_{n\to\infty} p_{ii}^{(n)}=\frac{1}{\mu_i}$$

如果 i 是周期为 t 的正常返状态，则

$$\lim_{n\to\infty} p_{ii}^{(nt)}=\frac{t}{\mu_i}$$

根据上述的定理 4, 5, 6, 可得如下的依赖于 $p_{ii}^{(n)}$ 的渐近性质的关于状态性质的判别法:

状态 i 非常返 $\Longleftrightarrow \sum\limits_{n=1}^{\infty} p_{ii}^{(n)}<\infty$（当然有 $\lim\limits_{n\to\infty} p_{ii}^{(n)}=0$）;

状态 i 零常返 $\Longleftrightarrow \sum\limits_{n=1}^{\infty} p_{ii}^{(n)}=\infty$, 且 $\lim\limits_{n\to\infty} p_{ii}^{(n)}=0$;

状态 i 正常返 $\Longleftrightarrow \sum\limits_{n=1}^{\infty} p_{ii}^{(n)}=\infty$, 且 $\overline{\lim\limits_{n\to\infty}}\, p_{ii}^{(n)}>0$.

---

① 要知道定理证明的读者，可参看复旦大学概率论教研组编译的"排队论"（上海科学技术出版社，1961 年）一书的附录 I 和 II.

其中$\overline{\lim\limits_{n\to\infty}}$表示上极限，特别在非周期的场合，必有

$$\overline{\lim_{n\to\infty}} p_{ii}^{(n)}=\lim_{n\to\infty} p_{ii}^{(n)}=\frac{1}{\mu_i}>0$$

我们来证明

定理 7 如 $i\longleftrightarrow j$，则它们或同为非常返或同为常返；在后一场合，或同为零常返或同为正常返.

[证明] 由于 $i\longleftrightarrow j$，所以存在 $l\geqslant 1$ 和 $n\geqslant 1$ 使 $p_{ij}^{(l)}=\alpha>0$，和 $p_{ji}^{(n)}=\beta>0$. 仿(1)式，可得

$$p_{ii}^{(l+m+n)}\geqslant p_{ij}^{(l)}p_{jj}^{(m)}p_{ji}^{(n)}=\alpha\cdot\beta\cdot p_{jj}^{(m)} \tag{4}$$

$$p_{jj}^{(l+m+n)}\geqslant p_{ji}^{(n)}p_{ii}^{(m)}p_{ij}^{(l)}=\alpha\cdot\beta\cdot p_{ii}^{(m)} \tag{5}$$

由此可知，$p_{ii}^{(m)}$和 $p_{jj}^{(m)}$的渐近性质是一样的，于是根据上述判别法即可证明定理.

定理 8 如 $i\longleftrightarrow j$，则它们或同为非周期的或同为周期的；在后一场合有相同的周期 t.

[证明] 由 (4)，(5) 易得前一场合，现证后一场合. 因为 $i\longleftrightarrow j$，所以有 $l\geqslant 1$ 和 $n\geqslant 1$，使 $p_{ij}^{(l)}=\alpha>0$，和 $p_{ji}^{(n)}=\beta>0$，且有

$$p_{ii}^{(l+n)}\geqslant p_{ij}^{(l)}p_{ji}^{(n)}=\alpha\beta>0$$

因此，如 i 是周期 t 的，则 $l+n$ 必为 t 的倍数. 类似地，

$$p_{jj}^{(l+n)}\geqslant p_{ji}^{(n)}p_{ij}^{(l)}=\alpha\beta>0$$

因此，如 j 是周期 t' 的，则 $l+n$ 必为 t' 的倍数. 再由(5)式，得

$$p_{jj}^{(l+m+n)}\geqslant p_{ji}^{(n)}p_{ii}^{(m)}p_{ij}^{(l)}=\alpha\beta\cdot p_{ii}^{(m)}$$

如果 $p_{ii}^{(m)}>0$，则由上式可知 $p_{jj}^{(l+m+n)}>0$. 根据周期的定义，$l+m+n$ 应是 t' 的倍数，但 $l+n$ 是 t' 的倍数，所以 m 也应是 t' 的倍数. 这就是说，所有使 $p_{ii}^{(m)}>0$ 的 m 都是 t' 的倍数，即除 $m=t'$，$2t'$，$3t'$，…外，$p_{ii}^{(m)}=0$，于是，根据周期的定义可知，t 也必须是 t' 的倍数. 由于 t 和 t' 的位置是对称的，所以同样可证：t' 必须是 t 的倍数. 于是 $t=t'$，即有相同的周期.

二、闭集和状态空间的分解

由一些状态组成的集合 C 称为是闭的(或闭集)，如果对任意 $i\in C$ 和 $j\notin C$，有：$p_{ij}=0$. 由此，进一步可推出，对 $i\in C$, $j\notin C$,

$$p_{ij}^{(2)}=\sum_{k=0}^{\infty}p_{ik}p_{kj}=\sum_{k\in C}p_{ik}p_{kj}+\sum_{k\notin C}p_{ik}p_{kj}$$
$$=0+0=0$$

用归纳法，可以证明，

$$p_{ij}^{(n)}=0,\ i\in C,\ j\notin C$$

即对 $i\in C$, $j\notin C$, 自状态 i 出发，不能到达状态 j.

由此显然可得，对一切 $n\geqslant 1$ 和 $i\in C$, 有

$$\sum_{j\in C}p_{ij}^{(n)}=1$$

显然，整个状态空间构成一个闭集，这是较大的闭集. 另一方面，吸收状态也构成一个闭集，这是较小的闭集.

除整个状态空间外，没有别的闭集的马尔可夫链称为是**不可约的**.

定理 9 所有常返状态构成一个闭集.

[证明] 先证：如 i 是常返状态，且 $i\to j$, 则 $j\to i$, 即 i, j 相通. 事实上，如自 j 出发不能到达 i, 则由于 $i\to j$, 于是自 i 出发到达 j 后，即不能再返回 i, 这与 i 是常返状态，$f_{ii}=1$, 即自 i 出发后，以概率 1 返回 i 的假定相矛盾. 此矛盾即证明上述的结论为真.

现在根据定理 7 立即可知，如 i 为常返状态，且 $i\to j$, 则 j 必为常返状态，亦即自常返状态出发，只能到达常返状态，不能到非常返状态. 换言之，常返状态全体构成一个闭集.

今后用 C 表示由所有常返状态组成的闭集.

系 不可约马尔可夫链或者没有非常返状态或者没有常返状

态.

定理 10 在C中, 相通关系$i\longleftrightarrow j$是等价关系, 即具有下列三个性质:

(1) 自返性: $i\longleftrightarrow i$;

(2) 对称性: 如$i\longleftrightarrow j$, 则$j\longleftrightarrow i$;

(3) 传递性: 如$i\longleftrightarrow j$, $j\longleftrightarrow k$, 则$i\longleftrightarrow k$.

[证明] (1) 由于在C中, $f_{ii}=1>0$, 由定理 2 的系立即推出.

(2) 由定义直接推出.

(3) 是定理 1 系 2 的特例.

由于每一等价关系决定一个分类, 于是用上述相通关系可得:

定理 11 所有的常返状态可分为若干个互不相交的闭集$\{C_h\}$, 且有:

(1) C_h中任二状态相通;

(2) C_h中的任一状态和C_g中的任一状态, 在$h\neq g$时, 互不相通.

今后将称C_1, C_2, …为**基本常返闭集**.

基于上述一些定理, 我们知道, 整个状态空间I可分解成:

$$I=N+C_1+C_2+\cdots$$

其中N是所有非常返状态组成的集合, C_1, C_2, …都是由常返状态组成的闭集, 这些常返闭集内部是相互相通的, 但两个常返闭集之间是互不相通的.

如从某一非常返状态出发, 系统可能一直在非常返集N中, 也可能进入某个基本常返闭集; 一旦进入某个基本常返闭集后, 即一直停留在这个基本常返闭集中. 如从某一常返状态出发, 则系统就一直停留在这个状态所在的基本常返闭集之中。

系 在只有常返状态的不可约链中, 所有状态都是相通的.

关于系统进入基本常返闭集后的运动情形，我们有：

定理 12 每一周期为 t 的基本常返闭集 C_h，可以进一步分解成 t 个互不相交的子集：

$$C_h = C_h(0) + C_h(1) + \cdots + C_h(t-1)$$

且这些子集可适当安排，使有如下性质：自 $C_h(l)$ 中的任一状态，下一步必然到达 $C_h(l+1)$ 中的某一个状态，其中如 $l=t-1$，则 $l+1=t$ 解释为 0，即 $l+1 \equiv t \equiv 0 \pmod t$。

定理的证明从略①。

根据这一定理，系统进入基本常返闭集后的运动，当此常返闭集的周期为 t 时，应是在 t 个子集中依次作循环的运动：

$$C_h(l) \to C_h(l+1) \to \cdots \to C_h(t-1) \to C_h(0) \to C_h(1) \to \cdots$$

对于有限齐次马尔可夫链，我们还有如下几个更进一步的结果：

定理 13 有限齐次马尔可夫链的所有非常返状态组成的集合 N 不可能是闭集，亦即不管系统自什么状态出发，迟早要进入常返闭集。

［证明］ 在有限马尔可夫链的场合，N 当然只能是有限的．同时根据定理 3，系统只能有限次返回非常返状态，从而只能有限次返回 N，于是迟早将进入常返闭集．

系 不可约的有限齐次马尔可夫链的状态都是常返的．

定理 14 有限齐次马尔可夫链没有零常返状态．

［证明］ 设有某个零常返状态 i，它属于某个基本常返闭集 C_h，于是有

$$\sum_{j \in C_h} p_{ij}^{(n)} = 1$$

---

① 可参看：K. L. Chung, "Markov Chains with Stationary Transition Probabilities"一书的 I. 3 节。

由于 i 是零常返的，$j\in C_k$ 又与 i 相通，因此 $j\in C_k$ 也都是零常返的．再根据定理 5 的系可知，对一切零常返的 j，有

$$\lim_{n\to\infty} p_{ij}^{(n)}=0$$

于是在上面和式中令 $n\to\infty$，由于左边和式只含有限项，可得

$$\lim_{n\to\infty}\sum_{j\in C_k} p_{ij}^{(n)}=\sum_{j\in C_k}\lim_{n\to\infty} p_{ij}^{(n)}=0$$

而右边常数 1 不变，即得矛盾．此矛盾证明了定理．

结合定理 13 的系和定理 14，可得

定理 15 不可约的有限齐次马尔可夫链的状态都是正常返的．

为了帮助理解上述的定理，我们来考察第一章中的一些马尔可夫链的例子．

在无限制的随机游动中，前已指出，整个状态空间都是相通的．由于

$$p_{ii}^{(n)}=\begin{cases}\binom{n}{\frac{n}{2}}(pq)^{\frac{n}{2}}, & n\text{ 是偶数}\\ 0, & n\text{ 是奇数}\end{cases}$$

显然可知所有状态都是周期 2 的．

又因为

$$p_{ii}^{(2n)}\doteq\frac{(4pq)^n}{\sqrt{\pi n}}$$

所以当 $p\neq q$ 时，$4pq<1$，因而 $\sum p_{ii}^{(n)}<\infty$，于是所有的状态都是非常返的．但当 $p=q=\frac{1}{2}$ 时，$4pq=1$，

$$p_{ii}^{(2n)}\doteq\frac{1}{\sqrt{\pi n}}$$

于是 $\sum p_{ii}^{(n)}=\infty$，且 $p_{ii}^{(n)}\to 0$．因此，所有的状态都是零常返的．

再来考察带有一个吸收壁的随机游动．这时有一个吸收状态0，它是正常返的，非周期的，并且自己就构成一个基本常返闭集．由于其它状态{1, 2, ⋯}都可到达它，而一旦到达它后，即一直停留在那儿，从而自任意其它状态出发，都不可能以概率1迟早返回本身，即当 $i=1, 2, \cdots$ 时，$f_{ii}<1$．于是所有其它的状态都是非常返的．仿前易知，所有其它的状态也都是周期为2的．

在本例中，整个状态空间 $I=\{0, 1, 2, \cdots\}$ 可分解为 $N=\{1, 2, \cdots\}$ 和 $C=\{0\}$ 之后：$I=N+C$．由于自 N 中的状态可到达 C，而自 C 中的状态不能到达 N，所以是不相通的．整个链也就不再是不可约的了．

现在来看带有两个反射壁的随机游动．这是一个有限马尔可夫链．由于整个状态空间都是相通的，因而还是一个不可约链，由于 $p_{00}=q>0$，所以状态0是非周期的，而所有状态都相通，所以所有的状态都是非周期的．由定理15可知，所有的状态都是正常返的，因为又是非周期的，所以所有的状态都是遍历的．这种马尔可夫链也简称为不可约的遍历链．

其它的例子，请读者自行分析．

三、$p_{ij}^{(n)}$的渐近性质与平稳分布

在本段中，我们将说明，$p_{ij}^{(n)}$ 在 $n\to\infty$ 时的极限在一定意义下都存在，而且它的值与初始状态 i 无关，同时还给出极限分布的一些性质．

定理16 如 j 是非常返或零常返状态，则

$$\lim_{n\to\infty} p_{ij}^{(n)}=0$$

[证明] 当 j 是零常返时，即为定理5的系．当 j 为非常返时，根据定理4，同样有 $\lim\limits_{n\to\infty} p_{jj}^{(n)}=0$，因而可以仿定理5系的证法，证明定理．

定理 17 如 j 是非周期的正常返状态，即 j 是遍历状态，则

$$\lim_{n\to\infty} p_{ij}^{(n)} = \frac{1}{\mu_j}\cdot f_{ij}$$

［证明］ 根据定理 6，有

$$\lim_{n\to\infty} p_{jj}^{(n)} = \frac{1}{\mu_j} > 0$$

再由(2)式，得

$$\begin{aligned} p_{ij}^{(n)} &= \sum_{l=1}^{n} f_{ij}^{(l)} p_{jj}^{(n-l)} \\ &= \sum_{l=1}^{n'} f_{ij}^{(l)} p_{jj}^{(n-l)} + \sum_{l=n'+1}^{n} f_{ij}^{(l)} p_{jj}^{(n-l)} \end{aligned}$$

即

$$\begin{aligned} 0 \leqslant p_{ij}^{(n)} - \sum_{l=1}^{n'} f_{ij}^{(l)} p_{jj}^{(n-l)} \\ = \sum_{l=n'+1}^{n} f_{ij}^{(l)} p_{jj}^{(n-l)} \leqslant \sum_{l=n'+1}^{n} f_{ij}^{(l)} \end{aligned}$$

由此仿定理 5 的系的推理，即得

$$\lim_{n\to\infty} p_{ij}^{(n)} = \sum_{l=1}^{\infty} f_{ij}^{(l)} \cdot \frac{1}{\mu_j} = f_{ij}\cdot\frac{1}{\mu_j}$$

在这一定理中出现了 f_{ij}，它一般是不容易求得的，但在有的场合是简单的。

定理 18 如 j 是常返状态，且 $j\to i$，即 $f_{ji}>0$，则 $f_{ij}=1$.

［证明］ 如果 $f_{ij}<1$，则自状态 i 出发，经过任意有限步都不能到达状态 j 的概率为 $1-f_{ij}>0$. 又因为 $j\to i$，因而自状态 j 出发，经有限步不能返回状态 j 的概率 $1-f_{jj}$ 将是正的，即 $f_{jj}<1$. 这与 j 是常返的矛盾，此矛盾证明定理为真。

系 在不可约的常返链中，即在所有状态都是常返状态且相互相通的马尔可夫链中，一切的 $f_{ij}=1$.

对于不可约链，根据定理 9 的系可知，或者它的所有状态都是非常返的，这时称为**不可约的非常返链**；或者它的所有状态都是常返的，这时称为**不可约的常返链**. 下面的定理给出了上述两种可能的一个判别准则.

定理 19 不可约链是常返的充要条件是下列方程组没有一个非零的有界解

$$z_i=\sum_{j=1}^{\infty}p_{ij}z_j, \qquad i=1, 2, \cdots \tag{6}$$

[证明] 分三步进行. 第一步，设

$$y_i^{(n)}=P\{T_{i0}>n\}, n=1, 2, \cdots$$

显然它关于 n 是单调下降的，并且它是非负的，因此下列极限存在

$$\begin{aligned} y_i &=\lim_{n\to\infty}y_i^{(n)}=\lim_{n\to\infty}P\{T_{i0}>n\} \\ &=1-\lim_{n\to\infty}P\{T_{i0}\leqslant n\}=1-f_{i0} \end{aligned}$$

另一方面，我们有

$$\begin{aligned} y_i^{(1)} &=P\{T_{i0}>1\}=P\{X(1)\neq 0|X(0)=i\} \\ &=\sum_{j=1}^{\infty}p_{ij} \end{aligned}$$

一般的，有

$$\begin{aligned} y_i^{(n)} &=P\{T_{i0}>n\} \\ &=P\{X(1)\neq 0, X(2)\neq 0, \cdots, X(n)\neq 0|X(0)=i\} \\ &=\sum_{j=1}^{\infty}P\{X(1)=j, X(2)\neq 0, \cdots, X(n)\neq 0|X(0)=i\} \\ &=\sum_{j=1}^{\infty}P\{X(1)=j|X(0)=i\}\cdot \\ &\quad P\{x(2)\neq 0, \cdots, X(n)\neq 0|X(0)=i, X(1)=j\} \end{aligned}$$

利用过程是齐次马尔可夫的，我们有

$$P\{X(2)\neq 0, \cdots, X(n)\neq 0|X(0)=i, X(1)=j\}$$

$$
\begin{aligned}
&=P\{X(2)\neq 0, \cdots, X(n)\neq 0|X(1)=j\}\\
&=P\{X(1)\neq 0, \cdots, X(n-1)\neq 0|X(0)=j\}\\
&=P\{T_{j0}>n-1\}=y_j^{(n-1)}
\end{aligned}
$$

于是，有

$$
y_i^{(n)}=\sum_{j=1}^{\infty}p_{ij}y_j^{(n-1)}
$$

由于 $p_{ij}\geqslant 0$ 和 $\sum_{j=1}^{\infty}p_{ij}\leqslant 1$，因而关于 n 的极限与和式可交换，两边可取极限 $n\to\infty$，得

$$
y_i=\sum_{j=1}^{\infty}p_{ij}y_j
$$

第二步，我们来证明，上述的 $\{y_i=1-f_{i0}\}$，$i=1,2,\cdots$ 是方程组(6)的最大有界解，即如 $\{z_i\}$ 为任一满足方程，且有界 $|z_i|\leqslant 1$ 的解，则有 $|z_i|\leqslant y_i$，$i=1,2,\cdots$。（这里取 $|z_i|\leqslant 1$ 并不失一般性，事实上，由于方程是齐次的，即如 $\{z_i\}$ 满足方程，则 $\{Kz_i\}$ 也满足方程，取 $K^{-1}=\max\limits_{1\leqslant i\leqslant\infty}|z_i|<\infty$，则相应的 $\{Kz_i\}$ 即满足要求。）

事实上，我们有

$$
|z_i|\leqslant\sum_{j=1}^{\infty}p_{ij}|z_j|\leqslant\sum_{j=1}^{\infty}p_{ij}=y_i^{(1)}
$$

于是

$$
|z_i|\leqslant\sum_{j=1}^{\infty}p_{ij}|z_j|\leqslant\sum_{j=1}^{\infty}p_{ij}y_j^{(1)}=y_i^{(2)}
$$

一般有

$$
|z_i|\leqslant y_i^{(n)}
$$

再令 $n\to\infty$，得

$$
|z_i|\leqslant y_i=1-f_{i0}
$$

第三步，我们来证明定理，如链是常返的不可约链，则由上一

定理的系可知，所有的 $f_{i0}=1$，即 $y_i=0$，而根据上述第二步可知，方程组的一切解：$|z_i|\leqslant y_i=0$，即方程组没有非零的有界解．反之，如方程组(6)没有有界的非零解，于是它的解 $\{y_i\}$ 必须全为零，即 $y_i=1-f_{i0}=0$，亦即 $f_{i0}=1$，$i=1, 2, \cdots$．进一步，我们考虑 f_{00}，即自 0 出发，迟早返回 0 的概率，它应等于自 0 出发，经一步就返回 0 的概率 p_{00}，加上自 0 出发，第一步到某一状态 $i(i\neq 0)$，然后自 i 出发迟早返回 0 的概率之和，即

$$f_{00}=p_{00}+\sum_{i=1}^{\infty}p_{0i}f_{i0}$$

由于 $f_{i0}=1$，所以

$$f_{00}=p_{00}+\sum_{i=1}^{\infty}p_{0i}=\sum_{i=0}^{\infty}p_{0i}=1$$

即状态 0 是常返的，亦即链是常返的，定理证毕．

对于不可约的常返链，根据定理 9 的系和定理 7 可知，或者它的所有状态都是零常返的，这时称为**不可约的零常返链**；或者它的所有状态都是正常返的，这时称为**不可约的正常返链**，特别，如果它的状态还是非周期的，则称为**不可约的遍历链**．

为了给出上述两种可能的一个判别准则，还需要引入平稳分布的概念．一个概率分布 $\{v_j\}$，即 $v_j\geqslant 0$ 和 $\sum_{j=0}^{\infty}v_j=1$，称为马尔可夫链 $\{p_{ij}\}$ 的**平稳分布**，如有

$$v_j=\sum_{i=0}^{\infty}v_ip_{ij} \tag{7}$$

对于平稳分布，有

$$\begin{aligned}v_j&=\sum_i v_ip_{ij}=\sum_i\Big(\sum_k v_kp_{ki}\Big)p_{ij}\\&=\sum_k v_k\Big(\sum_i p_{ki}p_{ij}\Big)=\sum_k v_kp_{kj}^{(2)}\end{aligned}$$

一般有

$$v_j=\sum_{i=0}^{\infty}v_ip_{ij}^{(n)} \tag{8}$$

而且，如果马尔可夫链的初始分布

$$P\{X(0)=j\}=f_j$$

恰好是平稳分布，则对一切 n，$X(n)$ 的分布也是平稳分布，且正好就是 f_i。事实上，我们有

$$\begin{aligned}P\{X(n)=j\}&=\sum_{i=0}^{\infty}P\{X(0)=i\}P\{X(n)=j|X(0)=i\}\\&=\sum_{i=0}^{\infty}f_ip_{ij}^{(n)}=f_j\end{aligned}$$

与平稳分布相关的是所谓极限分布的概念。在不可约链的场合，由定理 17 和 18 可知，如 j 是非周期的，则

$$\lim_{n\to\infty}p_{ij}^{(n)}=\frac{1}{\mu_j}\geqslant 0$$

今后称 $\left\{\frac{1}{\mu_j}\right\}$ **为极限分布**。

我们有:

定理 20 非周期不可约链是正常返的充要条件是它存在平稳分布，且此时平稳分布就是极限分布。

［证明］ **充分性** 设存在平稳分布 $\{v_j\}$，于是由(8)得:

$$v_j=\sum_{i=0}^{\infty}v_ip_{ij}^{(n)}$$

由于 $v_i\geqslant 0$ 和 $\sum_{i=0}^{\infty}v_i=1$，于是极限号与和式可交换，两边令 $n\to\infty$ 可得

$$v_j=\lim_{n\to\infty}\sum_i v_ip_{ij}^{(n)}=\sum_i v_i\left(\frac{1}{\mu_j}\right)=\frac{1}{\mu_j}$$

因为$\sum\limits_{j=0}^{\infty} v_j=1$，于是至少有一个 $v_k>0$，即$\dfrac{1}{\mu_k}>0$，从而

$$\lim_{n\to\infty} p_{ik}^{(n)}=\frac{1}{\mu_k}>0$$

由此，根据定理 6 后面的说明，k 是正常返的，从而整个链是正常返的，而且所有的$\dfrac{1}{\mu_j}=v_j>0$.

必要性 设链是正常返的，于是

$$\lim_{n\to\infty} p_{ij}^{(n)}=\frac{1}{\mu_j}>0$$

另一方面，根据切普曼-柯尔莫哥洛夫方程，有

$$p_{ij}^{(m+n)}=\sum_{k=0}^{\infty} p_{ik}^{(m)} p_{kj}^{(n)} \geqslant \sum_{k=0}^{K} p_{jk}^{(m)} p_{kj}^{(n)}$$

两边令 $m\to\infty$，可得

$$\frac{1}{\mu_j} \geqslant \sum_{k=0}^{K}\left(\frac{1}{\mu_k}\right)\cdot p_{kj}^{(n)}$$

再令 $K\to\infty$，得

$$\frac{1}{\mu_j} \geqslant \sum_{k=0}^{\infty}\left(\frac{1}{\mu_k}\right) p_{kj}^{(n)} \tag{9}$$

进一步要证明，只能成立等式．由

$$1=\sum_{k=0}^{\infty} p_{ik}^{(n)} \geqslant \sum_{k=0}^{K} p_{ik}^{(n)}$$

先令 $n\to\infty$，再令 $K\to\infty$，可得

$$1 \geqslant \sum_{k=0}^{\infty}\left(\frac{1}{\mu_k}\right)$$

将(9)式对 j 相加，如(9)对某个 j 严格不等式成立，则有

$$1 \geqslant \sum_{j=0}^{\infty} \frac{1}{\mu_j} > \sum_{j=0}^{\infty}\left[\sum_{k=0}^{\infty}\left(\frac{1}{\mu_k}\right) p_{kj}^{(n)}\right]$$

$$=\sum_{k=0}^{\infty}\left[\frac{1}{\mu_k}\cdot\sum_{j=0}^{\infty}p_{kj}^{(n)}\right]$$

$$=\sum_{k=0}^{\infty}\frac{1}{\mu_k}$$

亦即如假定(9)式对某一j严格不等式成立，将导至

$$\sum_{j=0}^{\infty}\frac{1}{\mu_j}>\sum_{k=0}^{\infty}\frac{1}{\mu_k}$$

这是矛盾结果，此矛盾证明(9)式对一切j都只能成立等式

$$\frac{1}{\mu_j}=\sum_{k=0}^{\infty}\frac{1}{\mu_k}p_{kj}^{(n)} \tag{10}$$

再令$n\to\infty$，得

$$\frac{1}{\mu_j}=\sum_{k=0}^{\infty}\frac{1}{\mu_k}(\lim_{n\to\infty}p_{kj}^{(n)})$$

$$=\left(\sum_{k=0}^{\infty}\frac{1}{\mu_k}\right)\frac{1}{\mu_j}$$

由此推知：

$$\sum_{k=0}^{\infty}\frac{1}{\mu_k}=1$$

由(10)式和上式，即知极限分布$\left\{\frac{1}{\mu_j}\right\}$就是平稳分布.

系 非周期不可约正常返链的平稳分布是唯一的，就是极限分布.

下面我们用上述一些定理来分析第一章中带一个反射壁的随机游动. 显然，这一马尔可夫链是非周期不可约的. 为了决定它是否是常返的，根据定理19，要考察方程组(6)有否非零的有界解. 现在的方程组(6)是：

$$z_1=\sum_{j=1}^{\infty}p_{1j}z_j=pz_2$$

$$z_i = \sum_{j=1}^{\infty} p_{ij} z_j = p z_{i-1} + p z_{i+1}, \qquad i=2, 3, \cdots$$

以上两式可改写为:

$$z_2 - z_1 = \frac{q}{p} z_1$$

$$z_{i+1} - z_i = \frac{q}{p}(z_i - z_{i-1}), \qquad i=2, 3, \cdots$$

由此易得:

$$z_{i+1} - z_i = \left(\frac{q}{p}\right)^i \cdot z_1, \qquad i=2, 3, \cdots$$

进而有

$$z_i - z_1 = \left[\left(\frac{q}{p}\right)^{i-1} + \left(\frac{q}{p}\right)^{i-2} + \cdots + \frac{q}{p}\right] z_1, \qquad i=2, 3, \cdots$$

最后得

$$z_i = \frac{1 - \left(\frac{q}{p}\right)^i}{1 - \frac{q}{p}} \cdot z_1, \qquad i=1, 2, 3, \cdots$$

由此显然可见,如 $p>q$,则 z_i 是有界的,于是链是非常返的.在$p=q=\frac{1}{2}$时,有 $z_i = i \cdot z_1$ 是无界的,在 $p<q$ 时,z_i 也是无界的,所以,在 $p \leqslant q$ 时,链是常返的.

现在进一步分析 $p \leqslant q$ 时,链什么时应是正常返,什么时应是零常返的.根据定理 20,要考察方程组(7),看是否存在平稳分布.现在的方程组(7)是:

$$v_0 = \sum_{i=0}^{\infty} v_i p_{i0} = q v_0 + q v_1$$

$$v_j = \sum_{i=0}^{\infty} v_i p_{ij} = v_{j-1} \cdot p + v_{j+1} \cdot q$$

它们可改写为:

$$v_1=\frac{p}{q}v_0$$

$$v_{j+1}-v_j=\frac{p}{q}(v_j-v_{j-1}),\qquad j=1,2,3,\cdots$$

仿前可解得

$$v_j=\left(\frac{p}{q}\right)^j\cdot v_0,\qquad j=0,1,2,3,\cdots$$

只要 $v_0\geqslant 0$, 则一切 $v_j\geqslant 0$, 再由要求 $\sum_{j=0}^{\infty}v_j=1$, 即

$$\left(1+\frac{p}{q}+\left(\frac{p}{q}\right)^2+\cdots\right)v_0=1$$

可以定出 v_0, 但在 $p=q=\frac{1}{2}$, 即 $\frac{p}{q}=1$ 时, 上述括号中的级数发散, 于是不存在平稳分布, 即此时链是零常返的, 而在 $p<q$ 时, 易得:

$$v_0=\left(\frac{1}{1-\frac{p}{q}}\right)^{-1}=1-\frac{p}{q}>0$$

$$v_j=\left(1-\frac{p}{q}\right)\left(\frac{p}{q}\right)^j,\qquad j=1,2,3,\cdots$$

它是平稳分布, 于是此时链是正常返的, 又因是非周期的, 所以也是遍历的.

综上所述, 有

$p>q \iff$ 链是非常返的

$p=q \iff$ 链是零常返的

$p<q \iff$ 链是正常返的

四、某些应用

离散的分支过程 我们来继续研究例 1.3.8, 并将证明那儿提到的关于离散分支过程的结论.

容易看出, 它的状态空间 $I=\{0,1,2,\cdots\}$. 由于一旦达到状态

0, 即群体灭种后，就一直处于状态0, 所以状态 0 是一个吸收状态，从而有 $p_{00}=1$.

今后我们将假定 $p_0=P\{Y=0\}>0$, 即每个个体没有下一代的概率是正的. 因为在 $p_0=0$ 时，我们要研究的灭种问题是没有意义的，在 $p_0>0$ 的假定下，所有除 0 以外的状态都是非常返的，这是因为

$$\begin{aligned}p_{i0}&=P(Y_1+\cdots+Y_i=0\}\\&=P\{Y_1=0, Y_2=0, \cdots, Y_i=0\}\\&=P\{Y_1=0\}\cdot P\{Y_2=0\}\cdots P\{Y_i=0\}\\&=p_0^i>0\end{aligned}$$

即从所有非零状态 i 都可达到状态 0, 而状态 0 则为吸收状态，一旦到达后即永远留在这一状态而无法再返回状态 i. 换言之，自任一非零状态 i 出发，不能以概率 1 再返回状态 i, 即 $f_{ii}<1$, 于是 i 是非常返的.

也可用定理 18 来证明上述结论. 事实上，如设某一非零状态 i 是常返的，则如上所述有 $f_{i0}>0$, 于是按定理 18, 应有 $f_{0i}=1$. 但由于 0 是吸收状态，所以这是不可能的，于是所有的非零状态 i 都是非常返的.

下面我们来研究第一章中提到的"灭种"问题，即研究到达吸收状态 0 的概率. 群体在 n 代之前灭种的概率为

$$p_{i0}^{(n)}=P\{X(n)=0|X(0)=i\}$$

如定义

$$\zeta_n=p_{10}^{(n)}=P\{X(n)=0|X(0)=1\}$$

则根据我们关于各个体产生下一代个体个数是相互独立的假定，有

$$p_{i0}^{(n)}=(\zeta_n)^i$$

因此，我们只要研究 ζ_n 即可，容易看出，

$$\zeta_1 = P\{X(1)=0 \mid X(0)=1\} = P\{Y=0\} = p_0$$

$$\begin{aligned}
\zeta_{n+1} &= P\{X(n+1)=0 \mid X(0)=1\} \\
&= \sum_{k=0}^{\infty} P\{X(n+1)=0, X(1)=k \mid X(0)=1\} \\
&= \sum_{k=0}^{\infty} P\{X(1)=k \mid X(0)=1\} P\{X(n+1)=0 \mid X(1)=k\} \\
&= \sum_{k=0}^{\infty} p_k \cdot P\{X(n)=0 \mid X(0)=k\} \\
&= \sum_{k=0}^{\infty} p_k \cdot p_{k0}^{(n)} = \sum_{k=0}^{\infty} p_k \cdot (p_{10}^{(n)})^k \\
&= G(\zeta_n), \qquad n \geqslant 0
\end{aligned}$$

其中 $G(z)$ 即例 1. 3. 8 中提到的母函数. 由于诸 p_k 是非负的, 所以 $G(z)$ 在 $0 \leqslant z \leqslant 1$ 中是单调非降的, 因此 ζ_n 也是单调非降的, 于是有极限

$$\zeta = \lim_{n \to \infty} \zeta_n$$

它就是群体迟早要灭种的概率, 由于 $\zeta_{n+1} = G(\zeta_n)$, 取极限即得

$$\zeta = G(\zeta) \tag{11}$$

我们还可以证明 ζ 是上述方程(11)的最小正根. 设 z 是(11)的任一正根, 则因为 $G(z)$ 是单调非降的, 所以有

$$\zeta_1 = p_0 = G(0) \leqslant G(z) = z$$

$$\zeta_2 = G(\zeta_1) \leqslant G(z) = z$$

一般有

$$\zeta_{n+1} = G(\zeta_n) \leqslant G(z) \leqslant z$$

再取极限, 即得

$$\zeta \leqslant z$$

这就证明了

定理 21 离散分支过程的灭种概率 ζ 是方程 (11) 的最小正

根.

关于方程(11),我们有如下的引理:

引理 设 $G(z)=\sum_{k=0}^{\infty}p_k z^k$,其中 $p_k\geqslant 0$, $\sum_{k=0}^{\infty}p_k=1$, 再设 $p_0>0$ 和

$$0<m=\sum_{k=1}^{\infty}k\cdot p_k=G'(1)<\infty$$

则方程(11)有一个满足 $0<\zeta<1$ 的根的充要条件是 $m>1$.

[证明] 考虑函数

$$f(z)=\frac{G(z)}{z}=\frac{p_0}{z}+p_1+p_2z+\cdots,\ 0<z\leqslant 1$$

我们有

$$f'(z)=-\frac{p_0}{z^2}+0+p_2+2p_3z+3p_4z^2+\cdots$$

$$f''(z)=2\cdot\frac{p_0}{z^3}+0+0+2p_3+6p_4z+\cdots>0$$

以及

$$\begin{aligned}
&f(0)=\infty,\quad f(1)=1\\
&f'(0)=-\infty\\
&f'(1)=-p_0+p_2+2p_3+3p_4+\cdots\\
&\qquad=(p_1+2p_2+3p_3+4p_4+\cdots)\\
&\qquad\quad-(p_1+p_2+p_3+p_4+\cdots)-p_0\\
&\qquad=m-1
\end{aligned}$$

因此,由 $f''(z)>0$, 可知 $f'(z)$是单调上升的, 且在[0, 1]连续,再由 $f'(0)=-\infty$和 $f'(1)=m-1$ 可知,当 $m-1>0$,即 $m>1$ 时,$f'(z)$有一个根 z_0, $0<z_0<1$,于是 $f'(z)$在$(0, z_0)$中是负的, 而在$(z_0, 1)$中是正的,因而 $f(z)$在$(0, z_0)$中是单调下降的, 在$(z_0, 1)$中是单调上升的, 即在 $z=z_0$ 处, $f(z)$达到极小值. 但 $f(1)=1$,

所以$f(z_0)<1$, 因而在$(0, z_0)$中有一个ζ, 使$f(\zeta)=1$,即$\zeta=G(\zeta)$, $0<\zeta<1$. 充分性得证.

反之,如$m-1<0$, 即$m<1$, 则$f'(1)<0$, 于是$f'(z)$在$0<z\leqslant 1$中恒为负, 在$0<z<1$中$f(z)$是单调下降的. 由于$f(1)=1$, 所以在$0<z<1$中$f(z)>1$, 即$G(z)>z$, 从而方程(11)没有满足$0<z<1$的根,必要性证毕.

应用上述引理和定理21, 可得下列关于离散分支过程灭种概率的定理.

定理 22 设$m=\sum_{k=1}^{\infty}k\cdot p_k$是每个个体产生下一代个体的平均个数,假定$p_0>0$和$0<m<\infty$, 则

(1) 当$m\leqslant 1$时,群体以概率1迟早灭种,即$\zeta=1$;

(2) 当$m>1$时,群体以概率ζ^i迟早灭种,而以概率$1-\zeta^i$变为无穷, 其中ζ是方程(11)的最小正根, i是初始时群体的大小,即$X(0)=i$.

[证明] (1)与(2)的前一半是定理21和引理的直接推论.(2)的后一半是因为所有非零状态都是非常返的,所以根据定理16有

$$\lim_{n\to\infty}p_{ij}^{(n)}=0, \qquad j\neq 0$$

因此,在$n\to\infty$时,群体中包含的个体个数,它或者是零,或者是无穷. 定理于是证毕.

顺便指出, 当$m>1$且i很大时,ζ^i变得很小,群体要变为无穷的. 概率$1-\zeta^i$就十分接近于1.

§2. 纯不连续马尔可夫过程

本节和下节中将研究时间t连续变化的马尔可夫过程$X(t)$, $t\in T$, 为确定起见, 恒假定$T=[0, \infty)$, 状态空间$X=(-\infty, \infty)$.

我们主要研究两类特殊的马尔可夫过程. 本节中讨论纯不连续马尔可夫过程，此时系统处于某一状态中不变，直至某一瞬间状态发生跳跃而到达一个新的状态，此后一直停留于这个新状态中直到发生新的跳跃为止。下节中讨论扩散过程，此时系统的状态随着时间变化一直在作连续的变化. 更一般的同时具有上述两种变化的马尔可夫过程，本书将不作讨论.

一、定义与例子

用第一章第三节的记号，以 $F(s, x; t, y)$ 表示时刻 s 时系统处于状态 x，而在时刻 t 时系统处于 $(-\infty, y)$ 中的概率，即

$$F(s, x; t, y) = P\{X(t) < y \mid X(s) = x\}, \qquad t \geqslant s$$

它当然具有第一章中所说的一些性质:

(1) $F(s, x; s, y) = \eta(x, y) = \begin{cases} 1, & x < y \\ 0, & x \geqslant y \end{cases}$

今后还假定满足连续性条件

$$\lim_{s \to t-0} F(s, x; t, y) = \lim_{t \to s+0} F(s, x; t, y) = \eta(x, y)$$

(2) $F(s, x; t, y)$ 关于 x 是可测函数，关于 y 是分布函数，即它是 y 的单调非降函数，且 $F(s, x; t, -\infty) = 0$ 和 $F(s, x; t, +\infty) = 1$.

(3) 满足切普曼-柯尔莫哥洛夫方程:

$$F(s, x; t, y) = \int_{-\infty}^{\infty} d_z F(s, x; u, z) F(u, z; t, y)$$

所谓纯不连续过程，则还要求 $F(s, x; t, y)$ 满足:

在 $X(t) = x$ 时，系统在 $(t, t+\Delta t)$ 中将以概率 $1 - q(t, x)\Delta t + o(\Delta t)$ ① 留在此状态中，而以概率 $q(t, x)\Delta t + o(\Delta t)$ 发生跳跃; 如

① $o(\Delta t)$ 是一个记号，它表示关于 Δt 的一个高阶无穷小量，即 $\lim\limits_{\Delta t \to 0} \frac{o(\Delta t)}{\Delta t} = 0$. 下面的 $o(1)$ 表示无穷小量，即 $\lim\limits_{\Delta t \to 0} o(1) = 0$.

发生跳跃，则$X(t+\Delta t)$的分布由分布 $Q(t, x; y)+o(1)$ 给出. 总起来，有

$$F(t, x; t+\Delta t, y)$$
$$=(1-q(t, x)\Delta t)\eta(x, y)+q(t, x)\Delta t Q(t, x; y)+o(\Delta t) \tag{1}$$

其中 $q(t, x)$ 常称为**跳跃率函数或跳跃强度函数**. 它总是非负的. $Q(t, x; y)$是一条件分布函数，因此关于 y 它是单调非降的，且有

$$Q(t, x; -\infty)=0, \qquad Q(t, x; +\infty)=1$$

又因为 $Q(t, x; y)$是近似描写 $X(t)=x$ 且在$(t, t+\Delta t)$发生跳跃的条件下 $X(t+\Delta t)$的分布，所以还必须有 $Q(t, x; y)$关于 y 在 $y=x$ 连续.

现在来说明，第一章中引入的普阿松过程是纯不连续过程，根据那儿的定义，并利用$X(t+\Delta t)-X(t)$与 $X(t)$ 的独立性，我们有：

$$F(t, x; t+\Delta t, y)=P\{X(t+\Delta t)<y \mid X(t)=x\}$$
$$=\sum_{z=0}^{k} P\{X(t+\Delta t)=z \mid X(t)=x\}$$
$$=\sum_{z=0}^{k} P\{X(t+\Delta t)=z, X(t)=x \mid X(t)=x\}$$
$$=\sum_{z=0}^{k} P\{X(t+\Delta t)-X(t)=z-x \mid X(t)=x\}$$
$$=\sum_{z=0}^{k} P\{X(t+\Delta t)-X(t)=z-x\}$$

其中 k 表示满足条件"$l<y$"的最大的整数 l; x, z 等均为整数. 根据普阿松过程的定义，当 $z-x<0$ 时，$P\{X(t+\Delta t)-X(t)=z-x\}$ 都是零，因而可以不计，于是当 $y\leqslant x$ 时，所有右边和式中的项都是零，从而 $F(t, x; t+\Delta t, y)=0$，而当 $y>x$ 时，有

$$F(t,x;t+\Delta t,y)=P\{X(t+\Delta t)-X(t)=0\}$$
$$+P\{X(t+\Delta t)-X(t)=1\}$$
$$+\sum_{l=2}^{k-x}P\{X(t+\Delta t)-X(t)=l\}$$

其中第一项表示在 $(t,t+\Delta t)$ 中系统状态不变的概率，它按普阿松过程的定义是

$$P\{X(t+\Delta t)-X(t)=0\}=e^{-\lambda\Delta t}$$
$$=1-\lambda\Delta t+o(\Delta t)$$

第二项表示在 $(t,t+\Delta t)$ 中系统状态发生跳跃，且恰好变到状态 $x+1$ 的概率，它是

$$P\{X(t+\Delta t)-X(t)=1\}=e^{-\lambda\Delta t}\cdot\lambda\Delta t$$
$$=\lambda\Delta t+o(\Delta t)$$

第三项表示在 $(t,t+\Delta t)$ 中系统状态发生跳跃，且变到 $x+1$ 之外其它状态的概率，它是

$$\sum_{l=2}^{k-x}P\{X(t+\Delta t)-X(t)=l\}=\sum_{l=2}^{k-x}e^{-\lambda\Delta t}\frac{(\lambda\Delta t)^l}{l!}$$
$$\leqslant\sum_{l=2}^{\infty}e^{-\lambda\Delta t}\frac{(\lambda\Delta t)^l}{l!}$$
$$=e^{-\lambda\Delta t}\left(\sum_{l=0}^{\infty}\frac{(\lambda\Delta t)^l}{l!}-1-\lambda\Delta t\right)$$
$$=1-e^{-\lambda\Delta t}(1+\lambda\Delta t)$$
$$=1-(1-\lambda\Delta t+o(\Delta t))(1+\lambda\Delta t)$$
$$=1-1+o(\Delta t)=o(\Delta t)$$

综上所述可知，在 $X(t)=x$ 时，系统在 $(t,t+\Delta t)$ 中以概率 $1-\lambda\Delta t+o(\Delta t)$ 留在此状态中，而以概率 $\lambda\Delta t+o(\Delta t)$ 发生跳跃；在发生跳跃的条件下，系统以概率 $1+o(1)$ 进入状态 $x+1$，因为

$$P\{X(t+\Delta t)-X(t)=1\mid X(t+\Delta t)-X(t)\geqslant 1\}$$

$$=\frac{P\{X(t+\Delta t)-X(t)=1\}}{P\{X(t+\Delta t)-X(t)\geqslant 1\}}$$

$$=\frac{\lambda\Delta t+o(\Delta t)}{\lambda\Delta t+o(\Delta t)}=1+o(\Delta t)$$

于是可知，普阿松过程是纯不连续的马尔可夫过程，相应地有

$$q(t,x)=\lambda$$

$$Q(t,x;y)=\begin{cases}1, & y>x+1\\ 0, & \text{其它}\end{cases}$$

顺便指出，这时的 $q(t,x)$ 和 $Q(t,x;y)$ 均与 t 无关．一般当 $q(t,x)$ 和 $Q(t,x;y)$ 与 t 无关时，相应的纯不连续的马尔可夫过程称为**齐次的**．因此，普阿松过程也是齐次的纯不连续的马尔可夫过程．

更多的纯不连续的马尔可夫过程将在本节的后面给出．

二、柯尔莫哥洛夫-费勒(Feller)积微分方程

在本段中，我们将证明纯不连续马尔可夫过程的转移概率 $F(s,x;t,y)$ 满足一对称为**柯尔莫哥洛夫-费勒积微分方程**的方程．

定理 1 设 $X(t)$ 是纯不连续的马尔可夫过程，$F(s,x;t,y)$ 是它的转移概率，它满足(1)式．假定它的 $q(t,x)$ 是有限，非负和关于 t 是连续的，$Q(t,x;y)$ 关于 t 是连续的，则有

$$\frac{\partial}{\partial s}F(s,x;t,y)=q(s,x)F(s,x;t,y)-q(s,x)\int_{-\infty}^{\infty}d_zQ(s,x;z)F(s,z;t,y)\tag{2}$$

如进一步假定 $q(t,x)$ 有界，则还有

$$\frac{\partial}{\partial t}F(s,x;t,y)=-\int_{-\infty}^{y}q(t,z)d_zF(s,x;t,z)+\int_{-\infty}^{\infty}q(t,z)Q(t,z;y)d_zF(s,x;t,z)\tag{3}$$

［证明］ 先证(2)．由切普曼-柯尔莫哥洛夫方程得

$$F(s-\Delta s, x; t, y)=\int_{-\infty}^{\infty} d_z F(s-\Delta s, x; s, z) F(s, z; t, y)$$

由(1)得

$$\begin{aligned} F(s-\Delta s, x; s, z)=&(1-q(s-\Delta s, x)\Delta s)\eta(x, z) \\ &+q(s-\Delta s, x)\Delta s Q(s-\Delta s, x; z) \\ &+o(\Delta s) \end{aligned}$$

将它代入上式，得

$$\begin{aligned} &F(s-\Delta s, x; t, y) \\ &=\int_{-\infty}^{\infty}(1-q(s-\Delta s, x)\Delta s) d_z\eta(x, z) F(s, z; t, y) \\ &\quad+q(s-\Delta s, x)\Delta s \int_{-\infty}^{\infty} d_z Q(s-\Delta s, x; z) F(s, z; t, y) \\ &\quad+o(\Delta s) \end{aligned}$$

由于

$$\int_{-\infty}^{\infty} d_z \eta(x, z) F(s, z; t, y)=F(s, x; t, y)$$

所以有

$$\begin{aligned} &F(s-\Delta s, x; t, y) \\ &=F(s, x; t, y)-q(s-\Delta s, x)\Delta s F(s, x; t, y) \\ &\quad+q(s-\Delta s, x)\Delta s \int_{-\infty}^{\infty} d_z Q(s-\Delta s, x; z) F(s, z; t, y) \\ &\quad+o(\Delta s) \end{aligned}$$

将右边第一项移到左边，再用 $-\Delta s$ 除两边，得

$$\begin{aligned} &\frac{F(s-\Delta s, x; t, y)-F(s, x; t, y)}{-\Delta s} \\ &=q(s-\Delta s, x)\cdot F(s, x; t, y) \\ &\quad-q(s-\Delta s, x)\int_{-\infty}^{\infty} dz Q(s-\Delta s, x; z) F(s, z; t, y)+o(1) \qquad (4) \end{aligned}$$

两边令 $\Delta s\to 0$，则由于连续性假定 $q(s-\Delta s,x)\to q(s,x)$，$Q(s-\Delta s,x;z)\to Q(s,x;z)$，$o(1)\to 0$，于是有

$$\frac{\partial F(s,x;t,y)}{\partial s}=q(s,x)F(s,x;t,y)$$
$$-q(s,x)\int_{-\infty}^{\infty}d_zQ(s,x;z)F(s,z;t,y)$$

由于切普曼-柯尔莫哥洛夫方程中的 Δs 必须是正的，所以上式的左边只是左导数，为了证明右导数也满足上述方程，我们有

$$\begin{aligned}
&F(s,x;t,y)\\
&=\int_{-\infty}^{\infty}d_zF(s,x;s+\Delta s,z)F(s+\Delta s,z;t,y)\\
&=(1-q(s,x)\Delta s)\int_{-\infty}^{\infty}d_z\eta(x,z)F(s+\Delta s,z;t,y)\\
&\quad+q(s,x)\Delta s\int_{-\infty}^{\infty}d_zQ(s,x;z)F(s+\Delta s,z;t,y)\\
&\quad+o(\Delta s)
\end{aligned}\tag{5}$$

如能证明 $F(s,x;t,y)$ 关于 s 是连续的，则由上式仿前面的方法可证 F 对 s 的右导数也存在，且满足(2)，(2)式于是得证。

$F(s,x;t,y)$ 的关于 s 的连续性可由下列的估计获得，由(4)有：

$$\begin{aligned}
&|F(s-\Delta s,x;t,y)-F(s,x;t,y)|\\
&\leqslant|q(s-\Delta s,x)F(s,x;t,y)|\cdot\Delta s\\
&\quad+|q(s-\Delta s,x)|\cdot\Delta s\cdot\left|\int_{-\infty}^{\infty}d_zQ(s-\Delta s,x;z)F(s,z;t,y)\right|\\
&\quad+o(\Delta s)\\
&\leqslant|q(s-\Delta s,x)|\Delta s+|q(s-\Delta s,x)|\Delta s+o(\Delta s)
\end{aligned}$$

其中利用了 $0\leqslant F(s,x;t,z)\leqslant 1$ 这一事实．类似地，用(5)可得

$$\begin{aligned}
&|F(s,x;t,y)-F(s+\Delta s,x;t,y)|\\
&\leqslant|q(s,x)|\Delta s+|q(s,x)|\Delta s+o(\Delta s)
\end{aligned}$$

由于$q(s,x)$关于s是连续的，所以当Δs充分小时，$|q(s-\Delta s,x)|$是有界的，因而以上两式右边在$\Delta s\to 0$时，都是趋于零的，从而证明了$F(s,x;t,y)$关于s的连续性.

现在来证(3)．由切普曼-柯尔莫哥洛夫方程得

$$F(s,x;t+\Delta t,y)=\int_{-\infty}^{\infty}d_zF(s,x;t,z)\cdot F(t,z;t+\Delta t,y)$$

由(1)得

$$F(t,z;t+\Delta t,y)=(1-q(t,z)\Delta t)\eta(z,y)$$
$$+q(t,z)\Delta tQ(t,z;y)+o(\Delta t)$$

代入上式得

$$\begin{aligned}
&F(s,x;t+\Delta t,y)\\
&=\int_{-\infty}^{\infty}d_zF(s,x;t,z)(1-q(t,z)\Delta t)\eta(z,y)\\
&\quad+\int_{-\infty}^{\infty}d_zF(s,x;t,z)(q(t,z)\Delta tQ(t,z;y))+o(\Delta t)\\
&=F(s,x;t,y)-\int_{-\infty}^{y}q(t,z)\Delta td_zF(s,x;t,z)\\
&\quad+\int_{-\infty}^{\infty}q(t,z)\Delta tQ(t,z;y)d_zF(s,x;t,z)+o(\Delta t)
\end{aligned}$$

将右边第一项移到左边，并用Δt除两边，得

$$\begin{aligned}
&\frac{F(s,x;t+\Delta t,y)-F(s,x;t,y)}{\Delta t}\\
&=-\int_{-\infty}^{y}d_zq(t,z)F(s,x;t,z)\\
&\quad+\int_{-\infty}^{y}q(t,z)Q(t,z;y)d_zF(s,x;t,z)+o(1).
\end{aligned}$$

令$\Delta t\to 0$，即知F对t的右导数满足(3)，为证明左导数也满足(3)，我们有：

$$F(s,x;t,y)=\int_{-\infty}^{\infty}d_zF(s,x;t-\Delta t,z)F(t-\Delta t,z;t,y)$$

$$=\int_{-\infty}^{\infty}(1-q(t-\Delta t, z)\Delta t)\eta(z, y)d_zF(s, x; t-\Delta t, z)$$

$$+\int_{-\infty}^{\infty}q(t-\Delta t, z)\Delta tQ(t-\Delta t, z; y)d_zF(s, x; t-\Delta t, z)$$

$$+o(\Delta t)$$

$$=F(s, x; t-\Delta t, y)-\int_{-\infty}^{y}q(t-\Delta t, z)\Delta td_zF(s, x; t-\Delta t, z)$$

$$+\int_{-\infty}^{\infty}q(t-\Delta t, z)\Delta tQ(t-\Delta t, z; y)d_zF(s, x; t-\Delta t, z)$$

$$+o(\Delta t)$$

将右边第一项移到左边，再用 Δt 除两边，得：

$$\frac{F(s, x; t-\Delta t, y)-F(s, x; t, y)}{-\Delta t}$$

$$=-\int_{-\infty}^{y}q(t-\Delta t, z)d_zF(s, x; t-\Delta t, z)$$

$$+\int_{-\infty}^{\infty}q(t-\Delta t, z)Q(t-\Delta t, z; y)d_zF(s, x; t-\Delta t, z)$$

$$+o(1)$$

仿前可证 $F(s, x; t, y)$ 关于 t 是连续的. 再利用 $q(t, x)$ 和 $Q(t, x; y)$关于 t 是连续的，以及 $q(t, x)$的有界性，两边令 $\Delta t\to 0$，即得所证.

上述的方程(2)一般称为**后退方程**，(3)称为**前进方程**.

有了上述的定理之后，一个自然的问题是它的反问题：给定 $q(t, x)$和 $Q(t, x; y)$后，是否存在一个转移概率$F(s, x; t, y)$使(1)，(2)，(3)成立，如存在是否是唯一的.

在定理 1 的条件下，回答是肯定的，但相应的证明比较冗长，这儿就不介绍了.

三、齐次可数的场合

在齐次可数的场合，进行的研究最多，结果也比较深入，在这一段中，我们将介绍其中的一部分，由于齐次可数的假定，一般的转移概率就变为$\{p_{ij}(t),\ i, j=0, 1, 2, \cdots\}$，它们显然有

$$p_{ij}(t)\geqslant 0 \tag{6}$$

$$\sum_{j=0}^{\infty} p_{ij}(t)=1 \tag{7}$$

和切普曼-柯尔莫哥洛夫方程：

$$p_{ij}(s+t)=\sum_{k=0}^{\infty} p_{ik}(s)p_{kj}(t) \tag{8}$$

此外，一般还假定

$$\lim_{t\to 0} p_{ij}(t)=\delta_{ij}=\begin{cases}1, & i=j\\ 0, & i\neq j\end{cases} \tag{9}$$

(9)通常称为**连续性条件**.

与上一段一般的场合不同，在齐次可数的场合，只要从上述(6)——(9)这些很自然的条件出发，就可以推出$p_{ij}(t)$的可微性，以及柯尔莫哥洛夫方程等等。

我们首先来证明：

定理 2 对固定的 i, j，$p_{ij}(t)$是 t 的一致连续函数。

［证明］ 设 $h>0$，则

$$\begin{aligned}p_{ij}(t+h)-p_{ij}(t)&=\sum_{k=0}^{\infty} p_{ik}(h)p_{kj}(t)-p_{ij}(t)\\ &=p_{ii}(h)p_{ij}(t)-p_{ij}(t)+\sum_{k\neq i} p_{ik}(h)p_{kj}(t)\\ &=-(1-p_{ii}(h))p_{ij}(t)+\sum_{k\neq i} p_{ik}(h)p_{kj}(t)\end{aligned}$$

由此推出

$$p_{ij}(t+h)-p_{ij}(t)\geqslant -(1-p_{ii}(h))p_{ij}(t)\geqslant -(1-p_{ii}(h))$$

和

$$p_{ij}(t+h)-p_{ij}(t)\leqslant\sum_{k\neq i}p_{ik}(h)p_{kj}(t)\leqslant\sum_{k\neq i}p_{ik}(h)=1-p_{ii}(h)$$

因此

$$|p_{ij}(t+h)-p_{ij}(t)|\leqslant 1-p_{ii}(h)$$

对 $h<0$ 时,可得类似的不等式,一般有

$$|p_{ij}(t+h)-p_{ij}(t)|\leqslant 1-p_{ii}(|h|)\to 0\quad(h\to 0)\tag{10}$$

下面进一步讨论 $p_{ij}(t)$ 的可微性.

定理 3 下列极限总存在:

$$\lim_{h\to 0+}\frac{1-p_{ii}(h)}{h}\equiv q_i\leqslant\infty$$

且对一切 $t>0$,总有

$$\frac{1-p_{ii}(t)}{t}\leqslant q_i\tag{11}$$

[证明] 记

$$\varliminf_{h\to 0+}\frac{1-p_{ii}(h)}{h}=\tilde{q}_i$$

如 $\tilde{q}_i=\infty$,则 q_i 存在,且 $q_i=\infty$,定理于是得证.

现在假定 $\tilde{q}_i<\infty$,首先来证明对任意 $t>0$,都有

$$\frac{1-p_{ii}(t)}{t}\leqslant\tilde{q}_i$$

设 $\varepsilon>0$ 和 $t>0$ 是任意给定的数,然后选足够小的 $h>0$,使

$$\frac{1-p_{ii}(h)}{h}\leqslant\tilde{q}_i+\frac{\varepsilon}{2},\quad h\left(\tilde{q}_i+\frac{\varepsilon}{2}\right)<1$$

和当 $0\leqslant s\leqslant h$ 时,

$$1-p_{ii}(s)<\frac{\varepsilon t}{2}$$

即

$$p_{ii}(h)\geqslant 1-h\left(\tilde{q}_i+\frac{\varepsilon}{2}\right)>0$$

和

$$p_{ii}(s) \geqslant 1 - \frac{\varepsilon t}{2}$$

再令 $t=nh+s$，其中 n 是整数，$0 \leqslant s < h$，由(8)可得

$$\begin{aligned} p_{ii}(t+s) &= \sum_{k=0}^{\infty} p_{ik}(t) p_{ki}(s) \\ &\geqslant p_{ii}(t) p_{ii}(s) \end{aligned}$$

于是

$$\begin{aligned} p_{ii}(t) &\geqslant [p_{ii}(h)]^n \cdot p_{ii}(s) \\ &\geqslant \left[1-h\left(\tilde{q}_i + \frac{\varepsilon}{2}\right)\right]^n \cdot \left(1-\frac{\varepsilon t}{2}\right) \\ &\geqslant \left[1-nh\left(\tilde{q}_i + \frac{\varepsilon}{2}\right)\right]\left(1-\frac{\varepsilon t}{2}\right) \\ &\geqslant 1-\frac{\varepsilon t}{2} - t\left(\tilde{q}_i + \frac{\varepsilon}{2}\right) = 1 - t(\tilde{q}_i + \varepsilon) \end{aligned}$$

即

$$\frac{1-p_{ii}(t)}{t} \leqslant \tilde{q}_i + \varepsilon$$

由于 ε 是任意的，所以

$$\frac{1-p_{ii}(t)}{t} \leqslant \tilde{q}_i$$

取上极限，得

$$\overline{\lim_{t \to 0+}} \frac{1-p_{ii}(t)}{t} \leqslant \tilde{q}_i = \underline{\lim_{h \to 0+}} \frac{1-p_{ii}(h)}{h}$$

所以

$$\begin{aligned} \overline{\lim_{t \to 0+}} \frac{1-p_{ii}(t)}{t} &= \underline{\lim_{t \to 0+}} \frac{1-p_{ii}(t)}{t} \\ &= \lim_{t \to 0+} \frac{1-p_{ii}(t)}{t} = q_i = \tilde{q}_i \end{aligned}$$

定理 4 如 $q_i < \infty$，则对一切 $t > 0$ 和 j，$p'_{ij}(t)$ 存在且连续，

而且还满足下列关系:

$$p'_{ij}(t+s)=\sum_{k=0}^{\infty}p'_{ik}(t)p_{kj}(s),\ t>0,\ s>0 \tag{12}$$

$$\sum_{k=0}^{\infty}p'_{ik}(t)=0 \tag{13}$$

$$\sum_{k=0}^{\infty}|p'_{ik}(t)|\leqslant 2q_i \tag{14}$$

[证明]①记

$$\Delta_{ij}(t,t+s)=\frac{p_{ij}(t+s)-p_{ij}(t)}{s}$$

则由(8)和(11)有

$$\begin{aligned}\Delta_{ij}(t,t+s)&\geqslant\frac{p_{ii}(s)p_{ij}(t)-p_{ij}(t)}{s}\\&=p_{ij}(t)\frac{p_{ii}(s)-1}{s}\\&\geqslant -q_ip_{ij}(t)\end{aligned} \tag{15}$$

用M表示$\{0,1,2,\cdots\}$的一个任意的子集,将上式对M中的j相加,得

$$\sum_{j\in M}\Delta_{ij}(t,t+s)\geqslant -q_i\cdot\sum_{j\in M}p_{ij}(t)\geqslant -q_i$$

由于

$$\begin{aligned}&\sum_{j=0}^{\infty}\Delta_{ij}(t,t+s)\\&=\frac{1}{s}\left(\sum_{j=0}^{\infty}p_{ij}(t+s)-\sum_{j=0}^{\infty}p_{ij}(t)\right)\\&=\frac{1}{s}(1-1)=0\end{aligned}$$

所以

① 证明中用到实变函数论中的结果,不熟悉的读者可略去。

$$\sum_{j \notin M}\Delta_{ij}(t, t+s) = -\sum_{j \in M}\Delta_{ij}(t, t+s) \leqslant q_i$$

由于M是$\{0, 1, 2, \cdots\}$的任意子集，所以如取M为其补集，则上式也成立，于是应当有

$$\left|\sum_{j \in M}\Delta_{ij}(t, t+s)\right| \leqslant q_i \tag{16}$$

现在令P是$\{0, 1, 2, \cdots\}$中使$\Delta_{ij}(t, t+s) \geqslant 0$的那些$j$的全体，于是由(16)，有

$$\begin{aligned}
&\sum_{j=0}^{\infty}|\Delta_{ij}(t, t+s)| \\
&= \sum_{j \in P}\Delta_{ij}(t, t+s) - \sum_{j \notin P}\Delta_{ij}(t, t+s) \\
&= \left|\sum_{j \in P}\Delta_{ij}(t, t+s)\right| + \left|\sum_{j \notin P}\Delta_{ij}(t, t+s)\right| \\
&\leqslant q_i + q_i = 2q_i
\end{aligned} \tag{17}$$

特别$|\Delta_{ij}(t, t+s)| \leqslant 2q_i$，亦即$p_{ij}(t)$满足李卜西兹条件，从而它是绝对连续的，因而由(9)，有

$$p_{ij}(t) = \delta_{ij} + \int_0^t p'_{ij}(u)du$$

其中积分是勒贝格积分，$p'_{ij}(u)$是关于勒贝格测度几乎处处存在，因此在(17)式两边取极限$s \to 0$，可知几乎处处有

$$\sum_{j=0}^{\infty}|p'_{ij}(t)| \leqslant 2q_i \tag{18}$$

因此可以逐项积分，

$$\begin{aligned}
&\int_0^t\left(\sum_{j=0}^{\infty}p'_{ij}(u)\right)du \\
&= \sum_{j=0}^{\infty}\int_0^t p'_{ij}(u)du = \sum_{j=0}^{\infty}(p_{ij}(t) - \delta_{ij})
\end{aligned}$$

$$=1-1=0$$

于是几乎处处有

$$\sum_{j=0}^{\infty} p'_{ij}(t)=0$$

由于 $p_{ij}(t)$ 是绝对连续的，所以对任何非负的 s, t 都有

$$p_{ij}(s+t)-p_{ij}(s)=\int_s^{s+t} p'_{ij}(u)du=\int_0^t p'_{ij}(s+u)du \tag{19}$$

另一方面，由(8)及(18)

$$\begin{aligned}p_{ij}(s+t)-p_{ij}(s)&=\sum_{k=0}^{\infty}(p_{ik}(t)-\delta_{ik})p_{kj}(s)\\&=\sum_{k=0}^{\infty}\int_0^t p'_{ik}(u)du p_{kj}(s)\\&=\int_0^t\sum_{k=0}^{\infty}p'_{ik}(u)p_{kj}(s)du\end{aligned}$$

比较上述两式可推得对每个 s，关于 t 几乎处处成立

$$p'_{ij}(t+s)=\sum_{k=0}^{\infty}p'_{ik}(t)p_{kj}(s)$$

至此，(12)——(14)关于 t 几乎处处成立．由于使(19)成立的 $p'_{ij}(t)$ 可以在零集上任意地改变，同时由(18)可推得(12)右端是 s 的连续函数，t 又可取得任意小，因而存在 $p'_{ij}(t)$ 的一种取法，它是 t 的连续函数，我们就规定 $p'_{ij}(t)$ 取这一连续函数，这时(13)，(14)仍关于 t 几乎处处成立，而(12)是对几乎所有的 t 对每个 s 成立．下面我们来说明(12)——(14)关于 t"几乎处处"成立的结论可推进为"处处"成立．对任一 $t>0$，取 τ 满足 $0<\tau<t$，且(12)当 $t=\tau$ 时对每个 s 成立，这时

$$\begin{aligned}p'_{ij}(t+s)&=p'_{ij}(\tau+(t+s-\tau))\\&=\sum_{l=0}^{\infty}p'_{il}(\tau)p_{lj}(t+s-\tau)\end{aligned}$$

$$=\sum_{l=0}^{\infty}p'_{il}(\tau)\sum_{k=0}^{\infty}p_{lk}(t-\tau)p_{kj}(s)$$

$$=\sum_{k=0}^{\infty}\left(\sum_{l=0}^{\infty}p'_{il}(\tau)p_{lk}(t-\tau)\right)p_{kj}(s)$$

$$=\sum_{k=0}^{\infty}p'_{ik}(t)p_{kj}(s)$$

即(12)对一切 $s\geqslant 0, t>0$ 成立。

类似地，对任一 $t>0$，可取 τ，满足 $0<\tau<t$ 且 τ 使(13)，(14)成立，这时利用(12)式有

$$\sum_{j=0}^{\infty}p'_{ij}(t)=\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}p'_{ik}(\tau)p_{kj}(t-\tau)$$

$$=\sum_{k=0}^{\infty}p'_{ik}(\tau)\sum_{j=0}^{\infty}p_{kj}(t-\tau)$$

$$=\sum_{k=0}^{\infty}p'_{ik}(\tau)=0$$

和

$$|p'_{ij}(t)|\leqslant\sum_{k=0}^{\infty}|p'_{ik}(\tau)|p_{kj}(t-\tau)$$

$$\sum_{j=0}^{\infty}|p'_{ij}(t)|\leqslant\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}|p'_{ik}(\tau)|p_{kj}(t-\tau)$$

$$=\sum_{k=0}^{\infty}|p'_{ik}(\tau)|\sum_{j=0}^{\infty}p_{kj}(t-\tau)$$

$$=\sum_{k=0}^{\infty}|p'_{ik}(\tau)|\leqslant 2q_i$$

同样由于上述的 τ 可以选得任意的小，所以(13)，(14)两式也对一切 $t>0$ 都成立，定理证毕。

上面对 $q_i<\infty$ 证明了 $p_{ij}(t)$ 在 $t>0$ 处的可微性，可以证明当 $q_i=\infty$ 时，定理同样是成立的。下面进一步证明它在 $t=0$ 处也是

可微的.

定理 5 下列极限总存在

$$\lim_{t\to 0}\frac{p_{ij}(t)}{t}\equiv q_{ij}<\infty,\ i\neq j$$

[证明] 由于(9), 对任意的 $e>0$, 存在 T, 使当 $0\leqslant t\leqslant T$ 时,

$$1-p_{ii}(t)<e,\quad 1-p_{jj}(t)<e$$

我们先来证明, 对任意满足关系 $0<t-h<nh\leqslant t\leqslant T$(其中 n 为整数)的正数 h 和 t, 总成立

$$p_{ij}(h)\leqslant\frac{p_{ij}(t)}{n}\cdot\frac{1}{1-3e}$$

为此, 我们来考察 $0, h, 2h, \cdots, nh\leqslant t$ 时的马尔可夫过程状态的变化. 用 A_k 表示系统从状态 i 出发, 在 $t=kh$ 时首次到达状态 j 这一事件. 容易看出 A_k, $1\leqslant k\leqslant n$ 是互不相容的. 再用 B_k 表示系统从状态 i 出发, 在 $t=kh$ 时又回到状态 i, 但在其间从未到达过状态 j 这一事件. 记 $p_k=P(A_k)$, $p=\sum_{k=1}^{n}p_k$, 和 $Q_k=P(B_k)$.

我们有

$$e>1-p_{ii}(t)=\sum_{k\neq i}p_{ik}(t)\geqslant p_{ij}(t)$$

$$\geqslant\sum_{k=1}^{n}p_k p_{jj}(t-kh)\geqslant(1-e)\sum_{k=1}^{n}p_k=(1-e)P$$

因此

$$p\leqslant\frac{e}{1-e}$$

还有

$$Q_k\geqslant p_{ii}\ (k\cdot h)-\sum_{r=1}^{k-1}p_r$$

$$\geqslant p_{ii}\ (k\cdot h)-p\geqslant 1-e-\frac{e}{1-e}$$

于是

$$p_{ij}(t)\geqslant\sum_{k=1}^{n}Q_{k-1}p_{ij}(h)p_{jj}(t-kh)$$

$$\geqslant n\left(1-\varepsilon-\frac{\varepsilon}{1-\varepsilon}\right)p_{ij}(h)(1-\varepsilon)$$

$$\geqslant n(1-3\varepsilon)p_{ij}(h)$$

即

$$p_{ij}(h)\leqslant\frac{p_{ij}(t)}{n}\cdot\frac{1}{1-3\varepsilon}$$

两边除以 h, 得

$$\frac{p_{ij}(h)}{h}\leqslant\frac{p_{ij}(t)}{nh}\cdot\frac{1}{1-3\varepsilon}\leqslant\frac{p_{ij}(t)}{t-h}\cdot\frac{1}{1-3\varepsilon}$$

先令 $h\to 0$, 得

$$\overline{\lim_{h\to 0+}}\frac{p_{ij}(h)}{h}\leqslant\frac{p_{ij}(t)}{t}\cdot\frac{1}{1-3\varepsilon}<\infty$$

再令 $t\to 0$, 得

$$\overline{\lim_{h\to 0+}}\frac{p_{ij}(h)}{h}\leqslant\left(\underline{\lim_{t\to 0+}}\frac{p_{ij}(t)}{t}\right)\cdot\frac{1}{1-3\varepsilon}<\infty$$

由于 ε 是任意的, 所以

$$\overline{\lim_{t\to 0+}}\frac{p_{ij}(t)}{t}=\underline{\lim_{t\to 0+}}\frac{p_{ij}(t)}{t}=\lim_{t\to 0+}\frac{p_{ij}(t)}{t}\equiv q_{ij}<\infty$$

系 对有限马尔可夫过程, $q_i<\infty$, 且

$$\infty>q_i=\sum_{j\neq i}q_{ij} \tag{20}$$

[证明] 由

$$1-p_{ii}(t)=\sum_{j\neq i}p_{ij}(t)$$

两边除以 t, 由于和式是有限的, 令 $t\to 0$, 引用定理即得所证.

但在状态无限时, (20)式一般不一定成立, 而只有

$$\sum_{j\neq i} q_{ij} \leqslant q_i$$

当(20)式成立时，称为**保守的**，上述的系 1 说明有限马尔可夫过程总是保守的.

下面进一步来讨论柯尔莫哥洛夫方程. 我们有：

定理 6 对给定的 i, $p_{ij}(t)$满足下列后退柯尔莫哥洛夫方程

$$p'_{ij}(t) = -q_i p_{ij}(t) + \sum_{k\neq i} q_{ik} p_{kj}(t) \tag{21}$$

的充要条件是：

$$q_i < \infty, \quad q_i = \sum_{j\neq i} q_{ij}$$

［证明］ **必要性** $q_i = \infty$ 时，(21)无意义，所以 q_i 只能小于 ∞，此外在 $q_i < \infty$时，将(21)式对 j 相加，由定理 4 的(13)知左边为 0，而右边为

$$-q_i + \sum_{j\neq i} q_{ij} = 0$$

充分性 对 $h > 0$，有

$$\begin{aligned}&\frac{p_{ij}(t+h) - p_{ij}(t)}{h}\\ &\qquad = -\frac{1-p_{ii}(h)}{h} p_{ij}(t) + \sum_{k\neq i} \frac{p_{ik}(h)}{h} p_{kj}(t)\end{aligned}$$

即

$$\begin{aligned}&\left|\frac{p_{ij}(t+h) - p_{ij}(t)}{h} + \frac{1-p_{ii}(h)}{h} p_{ij}(t) - \sum_{\substack{k\neq i\\ k\leqslant K}} \frac{p_{ik}(h)}{h} p_{kj}(t)\right|\\ &\quad = \left|\sum_{\substack{k\neq i\\ k>K}} \frac{p_{ik}(h)}{h} p_{kj}(t)\right| \leqslant \sum_{\substack{k\neq i\\ k>K}} \frac{p_{ik}(h)}{h}\\ &\quad = \frac{1-p_{ii}(h)}{h} - \sum_{\substack{k\neq i\\ k\leqslant K}} \frac{p_{ik}(h)}{h}\end{aligned}$$

其中 K 为任一正整数，两边令 $h \to 0$，根据前面的定理有

$$|p'_{ij}(t)+q_ip_{ij}(t)-\sum_{\substack{k\neq i\\ k\leq K}}q_{ik}p_{kj}(t)|$$

$$\leq q_i-\sum_{\substack{k\neq i\\ k\leq K}}q_{ik}$$

由定理的条件，可知上式右边在 $K\to\infty$ 趋于零，于是有(21)。

定理 7 对给定的 j，如 $q_j<\infty$，且

$$\lim_{h\to 0}\frac{p_{kj}(h)}{h}=q_{kj}$$

关于 k 一致成立，则下列前进柯尔莫哥洛夫方程成立：

$$p'_{ij}(t)=-p_{ij}(t)q_j+\sum_{k\neq j}p_{ik}(t)q_{kj} \tag{22}$$

[证明] 很简单，留给读者.

我们还可以进一步求得无条件概率 $p_j(t)=P\{x(t)=j\}=\sum_{i=1}^{\infty}p_ip_{ij}(t)$ 所满足的微分方程，其中 $p_i=P\{x(0)=i\}$ 是初始分布.

为此用 p_i 乘(22)两边，再对 i 相加，可得

$$p'_j(t)=-p_j(t)q_j+\sum_{k\neq j}p_k(t)q_{kj} \tag{23}$$

特别如存在**平稳分布**，即如有 $p_j(t)=\pi_j=$常数，则上式左边为 0，于是有

$$-\pi_jq_j+\sum_{k\neq j}\pi_kq_{kj}=0$$

或者

$$\sum_{k\neq j}\pi_kq_{kj}=\pi_jq_j \tag{24}$$

这就是平稳分布所必需满足的方程.

关于柯尔莫哥洛夫方程解的存在性和唯一性等比较复杂，这儿不再讨论. 但在一些特定场合，它们是可以直接解出的；而在另

一些特定场合，虽不能直接解出它们，但却能直接解出(24)．这些我们将在下一段中继续进行．

上面我们讨论了 $p_{ij}(t)$ 在 $t\to 0$ 时的性质，即 $p_{ij}(t)$ 在 $t=0$ 处的连续性，可微性等等，这些可以说是 $p_{ij}(t)$ 的无穷小性质．下面再来简单地讨论一下，$p_{ij}(t)$ 在 $t\to\infty$ 时的性质，也就是 $p_{ij}(t)$ 的遍历性质，它与上面提到的平稳分布有着直接的关系．

为简化起见，我们假定所有状态都是相通的，即对任意的 i 和 j，存在 $s>0$ 和 $t>0$，使

$$p_{ij}(s)>0 \text{ 和 } p_{ji}(t)>0$$

对任一 $h>0$，显然 $\{p_{ij}(h)\}$ 是某个马尔可夫链的一步转移概率，且有

$$p_{ij}^{(n)}(h)=p_{ij}(nh)$$

我们将称 $\{p_{ij}(h)\}$ 是 $\{p_{ij}(t),\ t\geqslant 0\}$ 的**步长为 h 的离散骨架．**

我们有

定理 8 $\{p_{ij}(h)\}$ 是不可约，非周期的．

［证明］ 首先由

$$p_{ii}(t)\geqslant\left[p_{ii}\left(\frac{t}{n}\right)\right]^n$$

和(9)，$p_{ii}(t)\to 1\ (t\to 0)$，可知，$p_{ii}(t)$ 对一切 $t\geqslant 0$，恒为正的．特别 $p_{ii}(nh)=p_{ii}^{(n)}(h)$ 都是正的，因而是非周期的．再来证不可约性，即要证对任意 i,j，存在 n 使

$$p_{ij}^{(n)}(h)=p_{ij}(nh)>0$$

事实上，根据我们的假定，对任意的 i,j，存在 $s>0$，使

$$p_{ij}(s)>0$$

选 n 充分大，使 $nh>s$，则由于 $p_{ii}(t)$ 恒大于零，有

$$p_{ij}(nh)\geqslant p_{ii}(nh-s)\,p_{ij}(s)>0$$

不可约性于是证毕．

系 对一切 h 和 i, j, 下列极限存在,且与 i 无关:

$$\lim_{n\to\infty} p_{ij}^{(n)}(h) = \lim_{n\to\infty} p_{ij}(nh) = \pi_j(h)$$

[证明] 由定理和定理 1.16, 1.17, 1.18 推出。

定理 9 对一切 i, j, 下列极限存在,且与 i 无关:

$$\lim_{t\to\infty} p_{ij}(t) = \pi_j$$

[证明] 由定理 8 的系可知,对任意 $h>0$, 有

$$\lim_{n\to\infty} p_{ij}(nh) = \pi_j(h)$$

于是对任意的 $\varepsilon>0$, 存在 N, 使 $n, n'\geqslant N$ 时,有

$$|p_{ij}(nh) - p_{ij}(n'h)| < \frac{\varepsilon}{3}$$

又根据定理 2, $p_{ij}(t)$是一致连续的,所以可选 h 充分小,使 $p_{ij}(t)$ 在任一长为 h 的区间内的两点上的差的绝对值小于 $\varepsilon/3$. 于是当 h 这样选定后,只要 $t, t'>Nh$, 就有

$$|p_{ij}(t)-p_{ij}(t')| \leqslant |p_{ij}(t)-p_{ij}(nh)|$$
$$+|p_{ij}(nh)-p_{ij}(n'h)|+|p_{ij}(n'h)-p_{ij}(t')|$$

如果 nh 选为 $nh\leqslant t<(n+1)h$, $n'h$ 选为 $n'h\leqslant t'\leqslant(n'+1)h$, 则

$$|p_{ij}(t)-p_{ij}(t')| < \varepsilon$$

也就是 $\lim\limits_{t\to\infty} p_{ij}(t)$存在,其值与 i 无关由上一定理之系直接推出. 事实上,对一切 h, $\pi_j(h)$ 是常数 π_j.

定理 10 $\{\pi_j\}$有下列性质

(i) $\pi_j\geqslant 0$;

(ii) $\sum\limits_{j=0}^{\infty}\pi_j\leqslant 1$;

(iii) 对任意 $s>0$, 有 $\pi_j=\sum\limits_{k=0}^{\infty}\pi_k p_{kj}(s)$.

[证明] (i)是显然的. 对任意 N, 有

$$\sum_{j=0}^{N} p_{ij}(t) \leqslant 1$$

两边取 $t \to \infty$ 得

$$\sum_{j=0}^{N} \pi_j \leqslant 1$$

再取 $N \to \infty$, 得

$$\sum_{j=0}^{\infty} \pi_j \leqslant 1$$

(ii)得证. 由

$$p_{ij}(t+s) = \sum_{k=0}^{\infty} p_{ik}(t) p_{kj}(s)$$

两边取 $t \to \infty$, 得

$$\pi_j \geqslant \sum_{k=0}^{\infty} \pi_k p_{kj}(s)$$

将上式对 j 相加, 得

$$\sum_{j=0}^{\infty} \pi_j \geqslant \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \pi_k p_{kj}(s)$$

$$= \sum_{k=0}^{\infty} \pi_k \sum_{j=0}^{\infty} p_{kj}(s) = \sum_{k=0}^{\infty} \pi_k$$

所以实际上只能成立等式:

$$\pi_j = \sum_{k=0}^{\infty} \pi_k p_{kj}(s)$$

与马尔可夫链时一样, 当 $\pi_j > 0$ 时, 称状态 j 为**正常返的**. 由于总是非周期的, 所以也是遍历的. 在这时, 我们有:

定理 11 如对某个 $\pi_j > 0$, 则对一切 i, $\pi_i > 0$, 且有

(i) $\sum_{j=0}^{\infty} \pi_j = 1$;

(ii) $\sum_{k \neq j} \pi_k q_{kj} = \pi_j q_j$, 即满足平稳分布所满足的方程(24).

［证明］ 根据定理9，π_j 可理解为 $\pi_j(h)$，因此除(ii)外，均可由马尔可夫链的相应的定理推出. (ii)的证明比较复杂①：

但在有限状态的条件下，是容易由定理9推出的. 事实上，由定理10的(iii)有

$$\pi_j=\sum_{k=0}^{\infty}\pi_k p_{kj}(s)$$

$$=\pi_j p_{jj}(s)+\sum_{k\neq j}\pi_k p_{kj}(s)$$

将右边第一项移到左边，再用 s 除两边，得

$$\pi_j\frac{1-p_{jj}(s)}{s}=\sum_{k\neq j}\pi_k\frac{p_{kj}(s)}{s}$$

两边令 $s\to 0$，即得所证.

四、一些例子

［例1］ **随机信号** 考虑计算机中某个触发器，它可能有两种状态，记为"0"与"1". 假定触发器状态的变化构成一个齐次的马尔可夫过程，且有

$$p_{01}(h)=\lambda h+o(h)$$
$$p_{10}(h)=\mu h+o(h)$$

由此可知，有

$$q_0=\lim_{h\to 0}\frac{1-p_{00}(h)}{h}$$

$$=\lim_{h\to 0}\frac{p_{01}(h)}{h}=\lambda=q_{01}$$

类似地

$$q_1=q_{10}=\mu$$

于是相应的前进柯尔莫哥洛夫方程是

① 参见吴立德，"可数马尔可夫过程状态的分类"，数学学报，第15卷(1965)，第32—41页.

$$p'_{00}(t)=-\lambda p_{00}(t)+\mu p_{01}(t)$$

$$p'_{01}(t)=-\mu p_{01}(t)+\lambda p_{00}(t)$$

$$p'_{10}(t)=-\lambda p_{10}(t)+\mu p_{11}(t)$$

$$p'_{11}(t)=-\mu p_{11}(t)+\lambda p_{01}(t)$$

在初始条件 $p_{00}(0)=p_{11}(0)=1$，$p_{10}(0)=p_{01}(0)=0$ 下，可以解得

$$p_{00}(t)=\lambda_0 e^{-(\lambda+\mu)t}+\mu_0$$

$$p_{01}(t)=\lambda_0[1-e^{-(\lambda+\mu)t}]$$

$$p_{10}(t)=\mu_0[1-e^{-(\lambda+\mu)t}]$$

$$p_{11}(t)=\lambda_0+\mu_0 e^{-(\lambda+\mu)t}$$

其中

$$\lambda_0=\frac{\lambda}{\lambda+\mu},\qquad \mu_0=\frac{\mu}{\lambda+\mu}$$

在 $t\to\infty$ 时，有

$$\lim_{t\to\infty}p_{00}(t)=\mu_0,\quad \lim_{t\to\infty}p_{10}(t)=\mu_0$$

$$\lim_{t\to\infty}p_{01}(t)=\lambda_0,\quad \lim_{t\to\infty}p_{11}(t)=\lambda_0$$

由此可见 $p_{ij}(t)$在 $t\to\infty$ 时，其极限值存在且与 i 无关。

如果过程的初始分布就取为上述的极限，即

$$p_0=P\{X(0)=0\}=\mu_0,\quad p_1=P\{X(0)=1\}=\lambda_0$$

则有

$$\begin{aligned}p_0(t)&=p_0p_{00}(t)+p_1p_{10}(t)\\&=\mu_0[\lambda_0e^{-(\lambda+\mu)t}+\mu_0]+\lambda_0\mu_0[1-e^{-(\lambda+\mu)t}]\\&=\mu_0^2+\lambda_0\mu_0=\mu_0(\lambda_0+\mu_0)=\mu_0\end{aligned}$$

类似地

$$p_1(t)=p_0p_{01}(t)+p_1p_{11}(t)=\lambda_0$$

也就是说，$p_0=\mu_0$，$p_1=\lambda_0$ 是平稳分布．在平稳态时，

$$m=EX(t)=0\cdot p_0+1\cdot p_1=p_1=\lambda_0$$

$$B(\tau)=E\{(X(t+\tau)-m)(X(t)-m)\}$$

$$=E\{X(t+\tau)X(t)\}-m^2$$
$$=p_1(t)p_{11}(\tau)-m^2$$
$$=\lambda_0[\lambda_0+\mu_0 e^{-(\lambda+\mu)\tau}]-\lambda_0^2$$
$$=\lambda_0\mu_0 e^{-(\lambda+\mu)\tau}$$

［例 2］ **生灭过程** 设 $X(t)$, $t\geqslant 0$ 是齐次可数马尔可夫过程, 如它的转移概率 $p_{ij}(t)$满足

$$p_{ii+1}(\tau)=\lambda_i\tau+o(\tau),\quad (\lambda_i>0)$$
$$p_{ii-1}(\tau)=\mu_i\tau+o(\tau),\quad (\mu_i>0,\ \mu_0=0)$$
$$p_{ii}(\tau)=1-(\lambda_i+\mu_i)\tau+o(\tau)$$
$$p_{ij}(\tau)=o(\tau),\quad |i-j|\geqslant 2$$

则称为生灭过程, 不难看出, 生灭过程的状态都是相通的.

上述诸式合起来有如下的概率解释: 在长为 τ 的一小段时间中, 在忽略高级无穷小项后, 只有三种可能, 状态由 i 变到 $i+1$, 也就是增加 1 (如将 $X(t)$理解为 t 时刻某群体的大小, 则就是生出一个个体), 其概率为 $\lambda_i\tau$; 由 i 变到 $i-1$, 也就是减少1, 或死去一个个体, 其概率为 $\mu_i\tau$; 或状态不变, 其概率为 $1-(\lambda_i+\mu_i)\tau$. 生灭过程的命名的理由也在于此.

不难看出, 相应的 q_i, q_{ij} 为.

$$q_i=\lambda_i+\mu_i$$
$$q_{ii-1}=\mu_i,\quad q_{ii+1}=\lambda_i$$
$$q_{ij}=0,\quad |i-j|\geqslant 2$$

而相应的后退柯尔莫哥洛夫方程为:

$$p'_{ij}(t)=-(\lambda_i+\mu_i)p_{ij}(t)+\lambda_i p_{i+1j}(t)+\mu_i p_{i-1j}(t)$$

前进柯尔莫哥洛夫方程为

$$p'_{ij}(t)=-p_{ij}(t)(\lambda_j+\mu_j)+p_{ij-1}(t)\lambda_{j-1}+p_{ij+1}(t)\mu_{j+1}$$

无条件概率满足的方程为

$$p'_j(t)=-p_j(t)(\lambda_j+\mu_j)+p_{j-1}(t)\lambda_{j-1}+p_{j+1}(t)\mu_{j+1}$$

$j=0, 1, 2, \cdots$

$$p_{-1}(t)=0$$

如假定平稳分布$\{p_i,\ i=0, 1, 2, \cdots\}$存在，则有：

$$-(\lambda_i+\mu_i)p_i+\lambda_{i-1}p_{i-1}+\mu_{i+1}p_{i+1}=0,\quad i=1, 2, \cdots$$

$$-\lambda_0 p_0+\mu_1 p_1=0$$

当一切$\mu_k>0$时，可逐步求得：

$$\begin{aligned} p_1&=\frac{\lambda_0}{\mu_1}p_0\\ p_2&=\frac{\lambda_0\lambda_1}{\mu_1\mu_2}p_0\\ &\cdots\cdots\cdots\cdots\\ p_k&=\frac{\lambda_0\lambda_1\cdots\lambda_{k-1}}{\mu_1\mu_2\cdots\mu_k}p_0\\ &\cdots\cdots\cdots\cdots \end{aligned} \tag{25}$$

再由$\sum_{k=0}^{\infty}p_k=1$，可知

$$p_0=\left(1+\sum_{k=1}^{\infty}\frac{\lambda_0\lambda_1\cdots\lambda_{k-1}}{\mu_1\mu_2\cdots\mu_k}\right)^{-1} \tag{26}$$

生灭过程是在生物、工程中大量出现的一种过程．下面是几个简单的例子．

机床维修 设有m台机床，s个维修工人($s\leqslant m$)．机床或者工作，或者损坏等待修理．机床损坏后，如有维修工人空着，则空着的工人立即来维修，否则等着，直到有一个工人修好另一部后再来修理，机床按先坏先修的原则进行排队．

进一步假定时刻t时正在工作的一台机床，在$(t, t+\Delta t)$中损坏的概率为$\lambda\Delta t+o(\Delta t)$；时刻$t$时正在修理的一台机床，在$(t, t+\Delta t)$中被修好的概率为$\mu\Delta t+o(\Delta t)$．并假定各机床之间的状态是相互独立的．

在上述假定下，如用 $X(t)$表示时刻 t 时损坏了的(包括正在维修和等待维修的，即不在工作的)机床个数，则可以看出它是一个时齐的有限的马尔可夫过程，$0 \leqslant X(t) \leqslant m$. 用 $p_{ij}(t)$记它的转移概率，根据上述假定，我们有:

$$p_{kk+1}(\Delta t) = (m-k)\lambda\Delta t + o(\Delta t), \quad k=0, 1, 2, \cdots, m-1$$

这是由于左边表示时刻 t 时有 k 台机床损坏，而在$(t, t+\Delta t)$中又有一台损坏的概率，在忽略掉一个高级无穷小后，它应等于在$(t, t+\Delta t)$中，原来正在工作着的 $m-k$ 台机床中恰好有一台损坏的概率，而后者正好就是$(m-k)\lambda\Delta t + o(\Delta t)$.

类似地，有

$$p_{kk-1}(\Delta t) = k\mu\Delta t + o(\Delta t), \quad 1 \leqslant k \leqslant s$$

$$p_{kk-1}(\Delta t) = s\mu\Delta t + o(\Delta t), \quad s \leqslant k \leqslant m$$

$$p_{kj}(\Delta t) = o(\Delta t), \ |k-j| \geqslant 2$$

于是它是一个生灭过程，相应的

$$\lambda_k = (m-k)\lambda, \quad k=0, 1, \cdots, m$$

$$\mu_k = \begin{cases} k\mu, & 1 \leqslant k \leqslant s \\ s\mu, & s \leqslant k \leqslant m \end{cases}$$

由(25)和(26)可知它的平稳分布是

$$\begin{aligned} p_k &= \frac{\lambda_0\lambda_1\cdots\lambda_{k-1}}{\mu_1\mu_2\cdots\mu_k}p_0 \\ &= \frac{m(m-1)\cdots(m-k+1)}{1\cdot 2\cdots k}\cdot\frac{\lambda^k}{\mu^k}p_0 \\ &= \binom{m}{k}\left(\frac{\lambda}{\mu}\right)^k p_0, \quad k \leqslant s \end{aligned}$$

$$\begin{aligned} p_k &= \frac{\lambda_0\cdots\lambda_{s-1}\lambda_s\cdots\lambda_k}{\mu_1\cdots\mu_s\mu_{s+1}\cdots\mu_k}p_0 \\ &= \frac{m(m-1)\cdots(m-s+1)(m-s)\cdots(m-k)}{1\cdot 2\cdots s\cdot s\cdots s}\cdot\frac{\lambda^k}{\mu^k}p_0 \end{aligned}$$

$$=\frac{m(m-1)\cdots(m-k)(s+1)(s+2)\cdots k}{1\cdot 2\cdots k\cdot s\cdot s\cdots s}\left(\frac{\lambda}{\mu}\right)^k p_0$$

$$=\binom{m}{k}\frac{(s+1)\cdots k}{s^{k-s}}\left(\frac{\lambda}{\mu}\right)^k p_0, \qquad s<k\leqslant m$$

$$p_0=\left[1+\sum_{k=0}^{s}\binom{m}{k}\left(\frac{\lambda}{\mu}\right)^k\right.$$

$$\left.+\sum_{k=s+1}^{m}\binom{m}{k}\frac{(s+1)(s+2)\cdots k}{s^{k-s}}\left(\frac{\lambda}{\mu}\right)^k\right]^{-1}$$

在给定了 m, λ, μ 之后，对于不同的 s，就可以用上述公式求出相应的$\{p_k\}$，进而求出相应的均值$\sum_{k=1}^{m}kp_k$(在安排 s 个维修工人时，平均不工作的机床数)等，根据这些数据以及增加维修工人的费用等即可用来决定合适的 s。

电话问题 考虑两个电话分局，假定它们之间有 s 条中继线路(中继线)，这两个电话局用户之间的通话要占用这些中继线路，每个电话局都有许多用户，其数量与 s 相比，一般都大得多，因此不管正在通话的用户有几个，不在通话的用户数几乎总可以当作是不变的，因此可以假定，在$(t, t+\Delta t)$中，又有一个用户要求通话的概率为 $\lambda\Delta t+o(\Delta t)$，而与正在通话的用户数无关；如此时有空着的电话线路，则上述用户就可以占用空着的中继线路而进行通话，否则上述用户的要求因线路占满而取消．再假定每一在时刻 t 正在通话的用户，在$(t, t+\Delta t)$中将结束通话，从而空出一条电话线路的概率为 $\mu\Delta t+o(\Delta t)$以及各用户之间是相互独立的．

在上述假定下，如用 $X(t)$表示时刻 t 时正在使用的中继线路的个数，则它是一个齐次的有限马尔可夫过程，记它的转移概率为 $p_{ij}(t)$，则有

$$p_{i\,i+1}(\Delta t)=\lambda\Delta t+o(\Delta t), \qquad i=0, 1, 2, \cdots, s-1$$

$$p_{i\,i-1}(\Delta t)=i\mu\Delta t+o(\Delta t),\quad i=1,2,\cdots,s$$

$$p_{ii}(\Delta t)=1-(\lambda+i\mu)\Delta t+o(\Delta t),\quad i=0,1,\cdots,s-1$$

$$p_{ss}(\Delta t)=1-s\mu\Delta t+o(\Delta t)$$

$$p_{ij}(\Delta t)=o(\Delta t),\quad |i-j|>1$$

于是它也是一个生灭过程，相应的

$$\lambda_k=\lambda,\qquad k=0,1,2,\cdots,s-1$$

$$\mu_k=k\mu,\qquad k=1,2,\cdots,s$$

由(25)和(26)可知它的平稳分布是

$$\begin{aligned}p_k&=\frac{\lambda_0\lambda_1\cdots\lambda_{k-1}}{\mu_1\mu_2\cdots\mu_k}p_0\\&=\frac{1}{k!}\left(\frac{\lambda}{\mu}\right)^k p_0,\qquad 1\leqslant k\leqslant s\end{aligned}$$

$$\begin{aligned}p_0&=\left[1+\sum_{k=1}^{s}\frac{1}{k!}\left(\frac{\lambda}{\mu}\right)^k\right]^{-1}\\&=\left[\sum_{k=0}^{s}\frac{1}{k!}\left(\frac{\lambda}{\mu}\right)^k\right]^{-1}\end{aligned}$$

即

$$p_k=\frac{\frac{1}{k!}\left(\frac{\lambda}{\mu}\right)^k}{\sum_{l=0}^{s}\frac{1}{l!}\left(\frac{\lambda}{\mu}\right)^l},\qquad 0\leqslant k\leqslant s \tag{27}$$

这就是有名的爱尔兰公式，它在电话工程中有广泛的应用。

§3. 扩散过程

一、定义和例子

本节中，我们将研究系统状态以连续方式变化的马尔可夫过程，由于历史上这类过程的研究起源于对扩散现象的研究，所以

就称为扩散过程.

假定过程在时刻 t 时位于 x, 即 $X(t)=x$, 我们来考察时间间隔$[t, t+\Delta t]$中的状态变化 $X(t+\Delta t)-X(t)$. 由于要求变化是连续的, 所以当 Δt 小的时候, $X(t+\Delta t)-X(t)$也应当是小的. 用式子来表示就是: 对任意 $\delta>0$, 有

$$\lim_{\Delta t\to 0}P\{|X(t+\Delta t)-X(t)|>\delta|X(t)=x\}=0 \tag{1}$$

或用转移概率 $F(s,x;t,y)=P\{X(t)<y|X(s)=x\}$写出

$$\lim_{\Delta t\to 0}\int_{|y-x|>\delta} d_yF(t,x;t+\Delta t,y)=0 \tag{1'}$$

一般我们要求较强的条件:

$$\lim_{\Delta t\to 0}\frac{1}{\Delta t}\int_{|y-x|>\delta} d_yF(t,x;t+\Delta t,y)=0 \tag{1''}$$

或更强的

$$\lim_{\Delta t\to 0}\frac{1}{\Delta t}\int_{|y-x|>\delta} (y-x)^2d_yF(t,x;t+\Delta t,y)=0 \tag{1'''}$$

除此而外, 我们还要求 $X(t+\Delta t)-X(t)$的均值与 Δt 成正比, 即

$$\lim_{\Delta t\to 0}\frac{1}{\Delta t}\int_{-\infty}^{\infty}(y-x)d_yF(t,x;t+\Delta t,y)=a(t,x) \tag{2}$$

以及 $X(t+\Delta t)-X(t)$的方差与 Δt 成正比, 即

$$\lim_{\Delta t\to 0}\frac{1}{\Delta t}\int_{-\infty}^{\infty}(y-x)^2d_yF(t,x;t+\Delta t,y)=b(t,x)>0 \tag{3}$$

特别指出, 上述极限均假定对 x 是一致的, 而且可以看到, 当条件(1''')成立时, 条件(2)和(3)分别可以改写为

$$\lim_{\Delta t\to 0}\frac{1}{\Delta t}\int_{|y-x|\leq\delta}(y-x)d_yF(t,x;t+\Delta t,y)=a(t,x) \tag{2'}$$

和

$$\lim_{\Delta t\to 0}\frac{1}{\Delta t}\int_{|y-x|\le\delta}(y-x)^2 d_y F(t,x,t+\Delta t,y)$$
$$=b(t,x)>0 \tag{3'}$$

数学上，一般称满足条件(1″)，(2′)，(3′)的马尔可夫过程为**扩散过程**；有时也称满足条件(1‴)，(2)，(3)的马尔可夫过程为**扩散过程**．显然，后一定义要求更强些．我们今后将用前一定义，定义中的 $a(t,x)$ 称为**偏移系数**，$b(t,x)$ 称为**扩散系数**，它总是非负的．

下面我们给出两个扩散过程的例子，它们都是由马尔可夫链经极限过渡而来的．这就告诉我们，在某些条件下，可以用连续的扩散过程来近似离散的过程，这种方法称为扩散近似，在有些复杂的离散过程的研究中是有用的．

［例 1］ **无限制的随机游动** 现在来考虑前面讨论过的无限制的随机游动，由例 1．3．1 可知它的 n 步转移概率是：

$$p_{ij}^{(n)}=\begin{cases}v(n,j-i), & \text{当 } n+j-i \text{ 为偶数}\\ 0, & \text{当 } n+j-i \text{ 为奇数}\end{cases}$$

其中

$$v(n,k)=\binom{n}{\frac{n+k}{2}}p^{\frac{n+k}{2}}q^{\frac{n-k}{2}},\quad \text{当 } n+k \text{ 为偶数}$$

进一步我们来计算 n 步后的位移，n 步产生的位移是由一步一步产生的位移累积起来的．每一步产生的位移可能是 $+1$，也可能是 -1，其概率分别为 p 和 q，因此每一步产生的平均位移是 $p-q$．于是 n 步产生的平均位移是 $n(p-q)$，同样每一步位移的方差是 $p+q-(p-q)^2=4pq$，又由于各步位移是相互独立的，因此 n 步后的位移的方差是 $4npq$．

为了研究极限过程，现在假定：(i)每步的空间长度为 Δx；(ii)每步所需的时间长度为 Δt．于是在 $(0,t)$ 这段时间中，总共产生

了 $t/\Delta t$ 步转移，相应的位移的均值和方差按上面所述应是：

$$\text{平均位移}=\frac{t}{\Delta t}\cdot(p-q)\cdot\Delta x=t\cdot(p-q)\cdot\frac{\Delta x}{\Delta t}$$

$$\text{位移方差}=4\cdot\frac{t}{\Delta t}\cdot p\cdot q\cdot(\Delta x)^2=4\cdot t\cdot p\cdot q\cdot\frac{(\Delta x)^2}{\Delta t}$$

由上式可知，为了使 $\Delta t\to 0$ 时极限方差有限，必须要求下列极限存在：

$$\lim_{\Delta t\to 0}\frac{(\Delta x)^2}{\Delta t}$$

而在这一条件下，要使平均位移也有限，则必须要求下列极限存在：

$$\lim_{\Delta t\to 0}\frac{p-q}{\Delta x}$$

为此，我们令

$$\frac{(\Delta x)^2}{\Delta t}=2D,\quad p=\frac{1}{2}+\frac{C}{2D}\Delta x,\quad q=\frac{1}{2}-\frac{C}{2D}\Delta x$$

其中 C 和 D 都是常数．这时容易看出前面所要求的两个极限一定都存在，事实上现在都是常数了，将它们代入平均位移和位移方差的表示式中，即得：

$(0,t)$ 中的平均位移 $\triangleq m(t)=2Ct$

$(0,t)$ 中的位移的方差 $\triangleq\sigma^2(t)=2Dt$

由此可见，位移的均值与方差都是与时间 t 成正比的，这也有助于理解上述定义中的条件（2）和(3)，且不难看出：$a(t,x)=2C$，$b(t,x)=2D$.

在 t 固定后，当 Δt 越来越小时，则 $n=t/\Delta t$ 越来越大，即转移的步数越来越大，于是 $(0,t)$ 中的位移可看成是越来越多的独立同分布的小位移之和，而其中每一小位移或取 Δx 或取 $-\Delta x$，相应的概率为 $p=\frac{1}{2}+\frac{C}{2D}\Delta x$ 和 $q=\frac{1}{2}-\frac{C}{2D}\Delta x$，因此应用中心极限定

理，可知在 $\Delta t\to 0$ 时，相应的二项分布 $v(n,k)$ 趋于一个正态分布。由于其均值和方差已经求出，所以其分布密度 $f(t,x)$ 应是：

$$f(t,x)=\frac{1}{\sqrt{2\pi}\,\sigma(x)}\exp\left\{-\frac{1}{2\cdot\sigma^2(t)}(x-m(t))^2\right\}$$

$$=\frac{1}{2\sqrt{\pi\cdot D\cdot t}}\exp\left\{-\frac{1}{4Dt}(x-2Ct)^2\right\}$$

它就是由无限制随机游动经上述极限过程所得的扩散过程在 $(0,t)$ 这段时间中所产生的位移的分布密度。如假定 $X(0)=0$，则它也就是上述扩散过程 $X(t)$ 的分布密度。

另一方面，由 $v(n,k)$ 的概率意义，即 $v(n,k)=p_{0k}^{(n)}$，易得：

$$p_{0k}^{(n+1)}=p_{0k-1}^{(n)}\cdot p_{k-1k}+p_{0k+1}^{(n)}\cdot p_{k+1k}$$

即
$$v(n+1,k)=p\cdot v(n,k-1)+q\cdot v(n,k+1)$$

考虑到时间和空间的步长 Δt 和 Δx，以及 $v(n,k)\to f(t,x)$，我们有

$$t=n\Delta t,\quad x=k\Delta x,\quad v(n,k)\doteq f(t,x)$$

从而上述递推式变为

$$f(t+\Delta t,x)=p\cdot f(t,x-\Delta x)+q\cdot f(t,x+\Delta x)$$

应用戴劳公式，并将 p,q 的表达式代入，可得

$$\begin{aligned}&f(t,x)+\frac{\partial f}{\partial t}\Delta t+o(\Delta t)\\&=\left(\frac{1}{2}+\frac{C}{2D}\Delta x\right)\left[f(t,x)-\frac{\partial f}{\partial x}\Delta x+\frac{\partial^2 f}{\partial x^2}\frac{(\Delta x)^2}{2}+o(\Delta x)^2\right]\\&\quad+\left(\frac{1}{2}-\frac{C}{2D}\Delta x\right)\left[f(t,x)+\frac{\partial f}{\partial x}\Delta x+\frac{\partial^2 f}{\partial x^2}\frac{(\Delta x)^2}{2}\right.\\&\quad\left.+o(\Delta x)^2\right]\end{aligned}$$

注意到 $\frac{(\Delta x)^2}{\Delta t}=o(1)$，上式可进一步化简为

$$\frac{\partial f}{\partial t}=-\frac{C}{D}\frac{\partial f}{\partial x}\frac{(\Delta x)^2}{\Delta t}+\frac{\partial^2 f}{\partial x^2}\frac{(\Delta x)^2}{2\Delta t}+o(1)$$

$$=-2C\frac{\partial f}{\partial x}+D\frac{\partial^2 f}{\partial x^2}+o(1)$$

当 $\Delta t\to 0$ 时，得 $f(t,x)$ 所满足的偏微分方程：

$$\frac{\partial f}{\partial t}=-2C\frac{\partial f}{\partial x}+D\frac{\partial^2 f}{\partial x^2}$$

其实对一般的扩散过程，也有类似的方程，这一点将在下一段中证明。

［例 2］ **爱伦菲斯特模型** 在第一章 § 3 第二段中曾讨论过这种模型，它的转移概率是：

$$p_{i\,i-1}=\frac{1}{2}\left(1+\frac{i}{a}\right),\ p_{i\,i+1}=\frac{1}{2}\left(1-\frac{i}{a}\right),\ -a<i<a$$

$$p_{a\,a-1}=1,\quad p_{-a\,-a+1}=1$$

由此可得

$$p_{ij}^{(n+1)}=p_{i\,j-1}^{(n)}p_{j-1\,j}+p_{i\,j+1}^{(n)}p_{j+1\,j}$$

$$=\frac{a-j+1}{2a}p_{i\,j-1}^{(n)}+\frac{a+j+1}{2a}p_{i\,j+1}^{(n)}$$

仿上例，选取时间步长 Δt 和空间步长 Δx，使

$$\frac{(\Delta x)^2}{\Delta t}=2D$$

随着 Δt 的减小，Δx 也减小，相应的 a 应增大，我们要求

$$a\cdot\Delta t=\beta^{-1}\qquad(\beta>0)$$

此时

$$a\cdot\Delta x=a\cdot(\Delta x)^2/\Delta x=a\cdot 2D\Delta t/\Delta x$$

$$=2D(a\Delta t)/\Delta x=2D/\beta\Delta x\to\infty$$

即随机游动的范围，随着 $\Delta t\to 0$ 而扩展到整个直线。

如假定 $t=0$ 时，系统在原点，并设 $f(t,x)$ 表示在 $\Delta t\to 0$ 时相应的随机游动的极限过程 $X(t)$ 的分布密度，于是如令

$$\frac{t}{\Delta t}=n,\quad \frac{x}{\Delta x}=j,\quad p_{0j}^{(n)}\doteq f(t,x)$$

则自上述差分方程可得：

$$f(t+\Delta t,x)=\frac{a-j+1}{2a}f(t,x-\Delta x)+\frac{a+j+1}{2a}f(t,x+\Delta x)$$

利用戴劳公式和关于 $\Delta x, \Delta t, a$ 的假定进行化简，可得

$$\begin{aligned}
&f(t,x)+\frac{\partial f}{\partial t}\Delta t+o(\Delta t)\\
&=\frac{a-j+1}{2a}\left[f(t,x)-\frac{\partial f}{2x}\Delta x+\frac{\partial^2 f}{\partial x^2}\frac{(\Delta x)^2}{2}+o(\Delta x)^2\right]\\
&+\frac{a+j+1}{2a}\left[f(t,x)+\frac{\partial f}{\partial x}\Delta x+\frac{\partial^2 f}{\partial x^2}\frac{(\Delta x)^2}{2}+o(\Delta x)^2\right]
\end{aligned}$$

或

$$\begin{aligned}
\frac{\partial f}{\partial t}&=\frac{1}{a\Delta t}f+\frac{j}{a\Delta t}\frac{\partial f}{\partial x}\Delta x+\left(1+\frac{1}{a}\right)\frac{\partial^2 f}{\partial x^2}\frac{(\Delta x)^2}{2\Delta t}+o(1)\\
&=\beta f+\beta x\frac{\partial f}{\partial x}+D\frac{\partial^2 f}{\partial x^2}+o(1)\\
&=\beta\frac{\partial(xf)}{\partial x}+D\frac{\partial^2 f}{\partial x^2}+o(1)
\end{aligned}$$

而在 $\Delta t\to 0$ 时，可得

$$\frac{\partial f}{\partial t}=\beta\frac{\partial(xf)}{\partial x}+D\frac{\partial^2 f}{\partial x^2}$$

二、柯尔莫哥洛夫方程

本段中将证明，满足上段(1″)(2′)(3′)三条件的马尔可夫过程，即扩散过程的转移函数 $F(s,x;t,y)$ 和转移密度

$$p(s,x;t,y)=\frac{\partial}{\partial y}F(s,x;t,y)$$

都满足一定的偏微分方程，它们称为**柯尔莫哥洛夫方程**，也称为**福克尔-普郎克(Fokker-Planck)方程**，这是因为这组方程在特殊的场合用不十分严格的方法，首先为福克尔和普郎克所得到，而在一般的场合，并用严格的方法，则为柯尔莫哥洛夫所得到。

定理1 设 $X(t)$, $t\in T=[a,b]$ 是扩散过程，并假定它的转移

概率密度函数 $p(s, x; t, y)$ 存在，且下列一些偏导数都存在并连续：

$$\frac{\partial}{\partial t}p(s,x;t,y)$$

$$\frac{\partial}{\partial y}[a(t,y)p(s,x;t,y)]$$

$$\frac{\partial^2}{\partial y^2}[b(t,y)p(s,x;t,y)]$$

则 $p(s,x;t,y)$ 满足下列方程

$$\frac{\partial}{\partial t}p(s,x;t,y) = -\frac{\partial}{\partial y}[a(t,y)p(s,x;t,y)] + \frac{1}{2}\frac{\partial^2}{\partial y^2}[b(t,y)p(s,x;t,y)]$$

［证明］ 设 $Q(y)$ 是有连续一阶、二阶导数的任意函数，且设 $Q(y)$ 在 $[a,b]$ 之外为零，则由 $Q(y), Q'(y), Q''(y)$ 的连续性，可知 $Q(y)$，$Q'(y)$ 和 $Q''(y)$ 在 $y=a$ 和 $y=b$ 处都是零．现在来考察下列积分：

$$\begin{aligned}
I &= \int_{-\infty}^{\infty} Q(y)\frac{\partial p}{\partial t}dy \\
&= \int_a^b Q(y)\frac{\partial}{\partial t}p(s,x;t,y)dy \\
&= \frac{\partial}{\partial t}\int_a^b Q(y)p(s,x;t,y)dy \\
&= \lim_{\Delta t\to 0}\int_a^b Q(y)\frac{p(s,x;t+\Delta t,y)-p(s,x;t,y)}{\Delta t}dy \\
&= \lim_{\Delta t\to 0}\frac{1}{\Delta t}\left\{\int_a^b Q(y)\left[\int_{-\infty}^{\infty} p(s,x;t,u)p(t,u;t+\Delta t,y)du\right]dy \right.\\
&\quad \left. -\int_a^b Q(y)p(s,x;t,y)dy\right\}
\end{aligned}$$

$$=\lim_{\Delta t\to 0}\frac{1}{\Delta t}\int_{-\infty}^{\infty}p(s,x;t,u)$$

$$\left[\int_a^b p(t,u;t+\Delta t,y)Q(y)dy-Q(u)\right]du$$

由于 $Q(y)$ 有界(闭区间上的连续函数必有界)和条件(1″)，我们有

$$\int_{|y-u|>\delta} d_yF(t,u;t+\Delta t,y)$$

$$=\int_{|y-u|>\delta} p(t,u;t+\Delta t,y)dy=o(\Delta t)$$

和

$$\int_{|y-u|>\delta} p(t,u;t+\Delta t,y)Q(y)dy=o(\Delta t)$$

于是

$$\frac{1}{\Delta t}\left[\int_a^b p(t,u;t+\Delta t,y)Q(y)dy-Q(u)\right]$$

$$=\frac{1}{\Delta t}\left[\int_{|y-u|\leqslant\delta} p(t,u;t+\Delta t,y)Q(y)dy-Q(u)\right]+o(1)$$

再在 u 的附近，将 $Q(y)$ 展开，得

$$Q(y)=Q(u)+Q'(u)(y-u)$$

$$+\frac{Q''(u)}{2}(y-u)^2+o((y-u)^2)$$

代入上式，得

$$\frac{1}{\Delta t}\left[\int_{|y-u|\leqslant\delta} p(t,u;t+\Delta t,y)Q(y)dy-Q(u)\right]$$

$$=\frac{1}{\Delta t}\left\{\int_{|y-u|\leqslant\delta} p(t,u;t+\Delta t,y)\right.$$

$$\cdot\left[Q(u)+Q'(u)(y-u)+\frac{Q''(u)}{2}(y-u)^2+o((y-u)^2)\right]dy$$

$$\left.-Q(u)\right\}$$

$$
=Q(u)\frac{1}{\Delta t}\int\limits_{|y-u|>\delta} p(t,u;t+\Delta t,y)dy
$$

$$
+Q'(u)\frac{1}{\Delta t}\int\limits_{|y-u|\leqslant\delta}(y-u)p(t,u;t+\Delta t,y)dy
$$

$$
+\frac{Q''(u)}{2}\left[\frac{1}{\Delta t}\int\limits_{|y-u|\leqslant\delta}(y-u)^2\cdot p(t,u;t+\Delta t,y)dy\right]
$$

$$
+o(1)
$$

应用条件(1″),(2′),(3′)可知在 $\Delta t\to 0$ 时,上式趋于

$$
Q'(u)a(t,u)+\frac{Q''(u)}{2}b(t,u)
$$

因此

$$
I=\int_{-\infty}^{\infty}p(s,x;t,u)\left[Q'(u)a(t,u)+\frac{Q''(u)}{2}b(t,u)\right]du
$$

$$
=\int_{-\infty}^{\infty}pQ'a\,du+\frac{1}{2}\int_{-\infty}^{\infty}pQ''b\,du
$$

利用分部积分法,有

$$
\int_{-\infty}^{\infty}apQ'du=apQ\Big|_{-\infty}^{\infty}-\int_{-\infty}^{\infty}Q\frac{\partial}{\partial u}(ap)du
$$

由于 $Q(u)$ 在 $[a,b]$ 之外为零,所以第一项为零,因此

$$
\int_{-\infty}^{\infty}apQ'du=-\int_{-\infty}^{\infty}\left[\frac{\partial}{\partial u}(ap)\right]\cdot Q du
$$

类似地

$$
\int_{-\infty}^{\infty}bpQ''du=\int_{-\infty}^{\infty}\left[\frac{\partial^2}{\partial u^2}(bp)\right]Q du
$$

所以

$$
I=\int_{-\infty}^{\infty}Q\frac{\partial p}{\partial t}du
$$

$$
=-\int_{-\infty}^{\infty}\left[\frac{\partial}{\partial u}(ap)\right]Q du+\frac{1}{2}\int_{-\infty}^{\infty}\left[\frac{\partial^2}{\partial u^2}(bp)\right]Q du
$$

$$=\int_{-\infty}^{\infty} Q\left[-\frac{\partial}{\partial u}(ap)+\frac{1}{2}\frac{\partial^2}{\partial u^2}(bp)\right]du$$

由于Q的任意性和$\frac{\partial p}{\partial t}$, $\frac{\partial}{\partial u}(ap)$, $\frac{\partial^2}{\partial u^2}(bp)$的连续性，即得

$$\frac{\partial}{\partial t}p(s,x;t,u)=-\frac{\partial}{\partial u}[a(t,u)p(s,x;t,u)]$$

$$+\frac{1}{2}\frac{\partial^2}{\partial u^2}[b(t,u)p(s,x;t,u)]$$

这正是我们所要证明的.

顺便指出，在上述证明过程中遇到的微分号与积分号交换以及积分次序的交换等，在我们现在的条件下，都是合理的.

除了上述的通常称为**前进方程**外，还有所谓的**后退方程**.

定理 2 设 $X(t)$, $t\in T=[a,b]$是扩散过程，再假定下列导数存在并连续：

$$\frac{\partial}{\partial x}F(s,x;t,y),\qquad \frac{\partial^2}{\partial x^2}F(s,x;t,y)$$

则 $F(s,x;t,y)$满足下列方程

$$\frac{\partial}{\partial s}F(s,x;t,y)=-a(s,x)\frac{\partial}{\partial x}F(s,x;t,y)$$

$$-\frac{1}{2}b(s,x)\frac{\partial^2}{\partial x^2}F(s,x;t,y)$$

[证明] 我们有

$$\frac{F(s-\Delta s,x;t,y)-F(s,x;t,y)}{\Delta s}$$

$$=\frac{1}{\Delta s}\left[\int_{-\infty}^{\infty} d_uF(s-\Delta s,x;s,u)F(s,u;t,y)\right.$$

$$\left.-F(s,x;t,y)\right]$$

$$=\frac{1}{\Delta s}\int_{-\infty}^{\infty} d_uF(s-\Delta s,x;s,u)[F(s,u;t,y)$$

$$-F(s,x;t,y)]$$

$$=I_1+I_2$$

其中

$$|I_1|=\left|\frac{1}{\Delta s}\int\limits_{|u-x|>\delta} d_uF(s-\Delta s,x;s,u)\right.$$

$$\left.\cdot[F(s,u;t,y)-F(s,x;t,y)]\right|$$

$$\leqslant\frac{2}{\Delta s}\int\limits_{|u-x|>\delta} d_uF(s-\Delta s,x;s,u)\to 0,\ (\Delta s\to 0)$$

$$I_2=\frac{1}{\Delta s}\int\limits_{|u-x|\leqslant\delta} d_uF(s-\Delta s,x;s,u)$$

$$\cdot[F(s,u;t,y)-F(s,x;t,y)]$$

$$=\frac{1}{\Delta s}\int\limits_{|u-x|\leqslant\delta} d_uF(s-\Delta s,x;s,u)$$

$$\cdot\left[\frac{\partial F(s,x;t,y)}{\partial x}(u-x)\right.$$

$$\left.+\frac{1}{2}\frac{\partial^2F(s,x;t,y)}{\partial x^2}(u-x)^2+o((u-x)^2)\right]$$

$$=\frac{\partial F(s,x;t,y)}{\partial x}\cdot\frac{1}{\Delta s}\int\limits_{|u-x|\leqslant\delta}(u-x)d_uF(s-\Delta s,x;s,u)$$

$$+\left(\frac{1}{2}\cdot\frac{\partial^2F(s,x;t,y)}{\partial x^2}+o(1)\right)$$

$$\cdot\frac{1}{\Delta s}\int\limits_{|u-x|\leqslant\delta}(u-x)^2d_uF(s-\Delta s,x;s,u)$$

$$\to a(s,x)\frac{\partial}{\partial x}F(s,x;t,y),\ +\frac{1}{2}b(s,x)\frac{\partial^2}{\partial x^2}F(s,x;t,y)$$

$$(\Delta s\to 0)$$

而当 $\Delta s\to 0$ 时，

$$\frac{F(s-\Delta s,x;t,y)-F(s,x;t,y)}{\Delta s}\to-\frac{\partial}{\partial s}F(s,x;t,y)$$

于是得到了所要证明的方程。

由于关于柯尔莫哥洛夫方程解的存在性与唯一性的讨论比较复杂，这儿就不进行了．下面我们考察两个简单的例子，在这种场合，柯尔莫哥洛夫方程的解是可以直接求出来的．

［例1］ **（维纳-爱因斯坦过程）** 这是一个齐次过程，它的转移概率密度 $p(t,x,y)$ 满足微分方程

$$\frac{\partial p}{\partial t}=D\frac{\partial^2 p}{\partial y^2}$$

其中 D 是一个正的常数．这相当于偏移系数 $a(t,x)=0$，扩散系数 $b(t,x)=2D$ 的扩散过程的前进柯尔莫哥洛夫方程．此外，还满足边界条件：

$$\lim_{y\to\pm\infty}p(t,x,y)=0,\quad \lim_{y\to\pm\infty}\frac{\partial}{\partial y}p(t,x,y)=0$$

不难看出，这一过程与上段例1在 $C=0$ 时（即 $p=q=\frac{1}{2}$ 时）是一致的．

现在来求解上述方程，引入 $p(t,x,y)$ 的特征函数，即

$$\phi(t,x,\theta)=\int_{-\infty}^{\infty}e^{i\theta y}p(t,x,y)dy$$

由于 $p(0,x,y)=\delta(x-y)$，所以

$$\begin{aligned}\phi(0,x,\theta)&=\int_{-\infty}^{\infty}e^{i\theta y}p(0,x,y)dy\\&=\int_{-\infty}^{\infty}e^{i\theta y}\delta(x-y)dy\\&=e^{i\theta x}\end{aligned}$$

应用上述边界条件和分部积分法，有

$$\begin{aligned}&\int_{-\infty}^{\infty}e^{i\theta y}\frac{\partial p}{\partial y}dy\\&=e^{i\theta y}p\Big|_{-\infty}^{\infty}-\int_{-\infty}^{\infty}p\cdot i\theta\cdot e^{i\theta y}dy\\&=0-i\theta\int_{-\infty}^{\infty}e^{i\theta y}p\cdot dy=-i\theta\phi(t,x,\theta)\end{aligned}$$

类似地

$$\int_{-\infty}^{\infty} e^{i\theta y}\frac{\partial^2 p}{\partial y^2}dy=(-i\theta)^2\phi(t,x,\theta)$$

$$=-\theta^2\cdot\phi(t,x,\theta)$$

用 $e^{i\theta y}$ 乘微分方程两边，并积分，得

$$\frac{\partial\phi(t,x,\theta)}{\partial t}=-D\theta^2\phi(t,x,\theta)$$

如固定 x,θ，这是关于 t 的一阶线性常微分方程，容易解得：

$$\phi(t,x,\theta)=Ae^{-D\theta^2 t}$$

其中 A 是常数，它可由 $\phi(0,x,\theta)=e^{i\theta x}$ 决定，即

$$A=\phi(0,x,\theta)=e^{i\theta x}$$

于是

$$\phi(t,x,\theta)=e^{i\theta x-D\theta^2 t}$$

$$=e^{i\theta x-\frac{1}{2}(2Dt)\theta^2}$$

容易看出，这是均值为 x，方差为 $2Dt$ 的正态分布的特征函数，因此

$$p(t,x,y)=\frac{1}{\sqrt{2\pi\cdot 2Dt}}\exp\left\{-\frac{1}{2\cdot 2Dt}(y-x)^2\right\}$$

$$=\frac{1}{2\sqrt{\pi Dt}}\exp\left\{-\frac{1}{4Dt}(y-x)^2\right\} \tag{4}$$

这也是维纳-爱因斯坦过程的转移密度．由上式可以看出，它关于空间也是齐次的，即 $p(t,x,y)$ 只依赖于 $y-x$，而与 x,y 本身的值无关，特别取 $x=0, y=x$，即从原点出发，在时间 t 内，到达 x 的概率密度便是

$$p(t,0,x)=\frac{1}{2\sqrt{\pi Dt}}\exp\left\{-\frac{x^2}{4Dt}\right\}$$

这与上段例 1 中，取 $C=0$ 时的结果是完全一致的．

下面我们进一步说明以(4)为转移概率密度的时齐马尔可夫

过程是扩散过程. 事实上, 由于

$$\frac{1}{\Delta t}\int_{-\infty}^{\infty}(x-y)^4 p(t,x,y)dy$$

$$=\frac{1}{\Delta t}\int_{-\infty}^{\infty}\frac{(x-y)^4}{\sqrt{4\pi D\Delta t}}\exp\left\{-\frac{(y-x)^2}{4D\Delta t}\right\}dy$$

$$=4D^2\Delta t\int_{-\infty}^{\infty}\frac{z^4}{\sqrt{2\pi}}e^{-\frac{z^2}{2}}dz=12D^2\Delta t$$

所以对固定的 δ 及 $i=0,1,2$ 有

$$\int_{|y-x|>\delta}(x-y)^i p(\Delta t,x,y)dy\leqslant\frac{1}{\delta^{4-i}}\int(x-y)^4 p(\Delta t,x,y)dy$$

$$=\frac{1}{\delta^{4-i}}12D^2(\Delta t)^2$$

又因

$$\int_{-\infty}^{\infty}(x-y)p(\Delta t,x,y)dy=0$$

$$\int_{-\infty}^{\infty}(x-y)^2 p(\Delta t,x,y)dy=2D\Delta t$$

所以(1‴), (2), (3)都对 x 一致地成立, 即以(4)为转移概率密度的马尔可夫过程为扩散过程, 其偏移系数 $a(t,x)\equiv 0$, 扩散系数 $b(t,x)=2D$. 进而若其初始分布满足

$$P\{X(0)=0\}=1$$

则此过程也就是维纳-爱因斯坦过程.

[例 2] (奥伦斯坦-乌伦贝克 (Ornstein-Uhlenbeck)过程) 这也是一个齐次的过程, 它的偏移系数和扩散系数分别是

$$a(t,x)=-\beta x,\quad(\beta>0)$$

$$b(t,x)=2\cdot D>0$$

相应的前进方程是

$$\frac{\partial p(t,x,y)}{\partial t}=\beta\frac{\partial}{\partial y}[y\cdot p(t,x,y)]+D\cdot\frac{\partial^2}{\partial y^2}p(t,x,y)$$

此外还满足边界条件

$$\lim_{y\to\pm\infty} yp(t,x,y)=0$$

$$\lim_{y\to\pm\infty}\frac{\partial}{\partial y}p(t,x,y)=0$$

不难看出，它与上段的例 2 是一致的.

现在来求解上述方程. 类似地引入特征函数，即

$$\phi(t,x,y)=\int_{-\infty}^{\infty}e^{i\theta y}p(t,x,y)dy$$

利用边界条件和分部积分法，同样可得

$$\begin{aligned}
&\int_{-\infty}^{\infty}e^{i\theta y}\frac{\partial}{\partial y}(yp)dy\\
&=e^{i\theta y}(yp)\Big|_{-\infty}^{\infty}-\int_{-\infty}^{\infty}yp\cdot i\theta\cdot e^{i\theta y}dy\\
&=0-\theta\cdot\int_{-\infty}^{\infty}p(iy)e^{i\theta y}dy\\
&=-\theta\int_{-\infty}^{\infty}p\cdot\frac{\partial(e^{i\theta y})}{\partial\theta}dy\\
&=-\theta\frac{\partial}{\partial\theta}\left[\int_{-\infty}^{\infty}e^{i\theta y}p\cdot dy\right]=-\theta\frac{\partial\phi}{\partial\theta}
\end{aligned}$$

和

$$\int_{-\infty}^{\infty}e^{i\theta y}\frac{\partial^2 p}{\partial y^2}dy=-\theta^2\phi(t,x,\theta)$$

仿上例，由微分方程可得

$$\frac{\partial\phi}{\partial t}=-\beta\theta\frac{\partial\phi}{\partial\theta}-D\theta^2\phi$$

为解上述一阶偏微分方程，我们来考虑相应的常微分方程

$$\frac{dt}{1}=\frac{d\theta}{\beta\theta}=\frac{d\phi}{-D\theta^2\phi}$$

解之可得两个首次积分：

$$\theta e^{-\beta t}=C_1$$

$$\phi\cdot\exp\left(\frac{D\theta^2}{2\beta}\right)=C_2$$

于是通解应为

$$\psi\left(\theta e^{-\beta t},\ \phi\cdot\exp\left(\frac{D\theta^2}{2\beta}\right)\right)=0$$

其中 ψ 是任意可微函数，或者对一个变元解出

$$\phi\cdot\exp\left(\frac{D\theta^2}{2\beta}\right)=\psi(\theta e^{-\beta t})$$

即

$$\phi(t,x,\theta)=\exp\left(-\frac{D\theta^2}{2\beta}\right)\cdot\psi(\theta e^{-\beta t})$$

其中 ψ 是任意可微函数，再由条件 $\phi(0,x,\theta)=e^{ix\theta}$，可得

$$\begin{aligned}\phi(0,x,\theta)&=\exp\left(-\frac{D\theta^2}{2\beta}\right)\psi(\theta)\\&=e^{ix\theta}\end{aligned}$$

即

$$\psi(\theta)=\exp\left(ix\theta+\frac{D\theta^2}{2\beta}\right)$$

于是最后求得

$$\begin{aligned}\phi(t,x,\theta)&=\exp\left(\frac{D\theta^2}{2\beta}\right)\psi(\theta e^{-\beta t})\\&=\exp\left(-\frac{D\theta^2}{2\beta}\right)\cdot\exp\left[ix\theta e^{-\beta t}+\frac{D}{2\beta}(\theta e^{-\beta t})^2\right]\\&=\exp\left[i(xe^{-\beta t})\,\theta-\frac{1}{2}\,\frac{D}{\beta}(1-e^{-2\beta t})\theta^2\right]\end{aligned}$$

容易看出这是均值为 $xe^{-\beta t}$，方差为 $\frac{D}{\beta}(1-e^{-2\beta t})$ 的正态分布的特征函数，因此

$$p(t,x,y)=\frac{1}{\sqrt{2\pi\frac{D}{\beta}(1-e^{-2\beta t})}}\cdot\exp\left[-\frac{1}{2\cdot\frac{D}{\beta}(1-e^{-2\beta t})}(y-xe^{-\beta t})^2\right]\tag{5}$$

类似于例 1，也可以验证具有上述转移概率密度的马尔可夫过程是满足(1‴)，(2)，(3)的，即它是扩散过程，而且相应于这一转移概率密度的均值与方差分别为

$$m(t)=xe^{-\beta t}$$

$$\sigma^2(t)=\frac{D}{\beta}(1-e^{-2\beta t})$$

容易看出，当 $t\to\infty$ 时，$m(t)\to 0$，$\sigma^2(t)\to\frac{D}{\beta}$，而相应的转移密度趋于正态分布：

$$\frac{1}{\sqrt{2\pi\frac{D}{\beta}}}\exp\left\{-\frac{y^2}{2\frac{D}{\beta}}\right\} \tag{6}$$

还可以看出，上述分布具有马尔可夫链中所述平稳分布的性质，即如果 $t=0$ 时，$X(0)$ 的分布是上述分布的话，则在任意时刻 t 时的分布也是这个分布。事实上，$X(t)$ 的分布密度是

$$f(t,y)=\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi\frac{D}{\beta}}}\exp\left[-\frac{x^2}{2\frac{D}{\beta}}\right]p(t,x,y)dx$$

$$=\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi\frac{D}{\beta}}}\exp\left[-\frac{x^2}{2\frac{D}{\beta}}\right]$$

$$\cdot\frac{1}{\sqrt{2\pi\frac{D}{\beta}(1-e^{-2\beta t})}}\exp\left[\frac{(y-xe^{-\beta t})^2}{2\frac{D}{\beta}(1-e^{-2\beta t})}\right]dx$$

$$=\frac{1}{\sqrt{2\pi\frac{D}{\beta}}\cdot\sqrt{2\pi\frac{D}{\beta}(1-e^{-2\beta t})}}$$

$$\cdot\int_{-\infty}^{\infty}\exp\left[\frac{x^2(1-e^{-2\beta t})+(y-xe^{-\beta t})^2}{2\frac{D}{\beta}(1-e^{-2\beta t})}\right]dx$$

$$=\frac{1}{\sqrt{2\pi\frac{D}{\beta}}\sqrt{2\pi\frac{D}{\beta}(1-e^{-2\beta t})}}$$

$$\cdot\int_{-\infty}^{\infty}\exp\left[-\frac{y^2(1-e^{-2\beta t})+(x-ye^{-\beta t})^2}{2\frac{D}{\beta}(1-e^{-2\beta t})}\right]dx$$

$$=\frac{1}{\sqrt{2\pi\frac{D}{\beta}}}\exp\left[-\frac{y^2}{2\cdot\frac{D}{\beta}}\right]\cdot\frac{1}{\sqrt{2\pi\frac{D}{\beta}(1-e^{-2\beta t})}}$$

$$\cdot\int_{-\infty}^{\infty}\exp\left[-\frac{(x-ye^{-\beta t})^2}{2\frac{D}{\beta}(1-e^{-2\beta t})}\right]dx$$

$$=\frac{1}{\sqrt{2\pi\frac{D}{\beta}}}\exp\left[-\frac{y^2}{2\frac{D}{\beta}}\right]$$

在上述推导中，我们顺便看出，如果 $X(0)$ 的分布是平稳分布，则 $X(0)$ 和 $X(t)$ 的联合分布密度是:

$$\frac{1}{\sqrt{2\pi\frac{D}{\beta}}}\exp\left[-\frac{x^2}{2\frac{D}{\beta}}\right]\cdot\frac{1}{\sqrt{2\pi\frac{D}{\beta}(1-e^{-2\beta t})}}$$

$$\cdot\exp\left[-\frac{(y-xe^{-\beta t})^2}{2\frac{D}{\beta}(1-e^{-2\beta t})}\right]$$

$$=\frac{1}{2\pi\frac{D}{\beta}\sqrt{1-e^{-2\beta t}}}\exp\left[-\frac{x^2-2e^{-\beta t}xy+y^2}{2\frac{D}{\beta}(1-e^{-2\beta t})}\right]$$

由此可知，$X(0)$ 和 $X(t)$（或一般地 $X(s)$ 和 $X(s+t)$）的协方差为

$$\operatorname{cov}(X(s),X(s+t))=\frac{D}{\beta}e^{-\beta|t|}\tag{7}$$

通常我们把以(6)为初始分布及以(5)为转移概率密度的马尔可夫过程称为**奥伦斯坦-乌伦贝克过程**. 所以(7)便是这一过程的协方差函数.

习　　题

1. 设 $U_1, U_2, \cdots, U_n$, 是相互独立的随机变量，试问下列的 $\{X_n, n\geq 1\}$ 是不是马尔可夫过程，并说明理由：

(i) $X_n = U_1 + U_2 + \cdots + U_n$;

(ii) $X_n = (U_1 + U_2 + \cdots + U_n)^2$;

(iii) $X_n = \rho X_{n-1} + U_n$, 其中 ρ 是一已知常数，$X_0 = 0$.

2. **一个满足切普曼-柯尔莫哥洛夫方程的非马尔可夫过程.** 设 $\{X_n\}$ 是如下定义的一串随机变量：考虑一串袋子，每一个袋子中包含四个球，分别编为 1, 2, 3, 4 号. 假定每次依次从一个袋子中取出一个球，对 $m=1, 2, \cdots$, 令

$A_m^{(1)}$ 表示"从第 m 个袋子中摸出的球是 1 号或 4 号"这一事件；

$A_m^{(2)}$ 表示"从第 m 个袋子中摸出的球是 2 号或 4 号"这一事件；

$A_m^{(3)}$ 表示"从第 m 个袋子中摸出的球是 3 号或 4 号"这一事件.

对 $m=1, 2, \cdots$ 和 $j=1, 2, 3$ 令

$$X_{3(m-1)+j} = \begin{cases} 1, & \text{如 } A_m^{(j)} \text{ 出现} \\ 0, & \text{如 } A_m^{(j)} \text{ 不出现} \end{cases}$$

试证：(i) 当 $n > m$ 时，对任意只取 0 和 1 的 k, k_1 和 k_2 有

$$P\{X_n = k\} = P\{X_n = k_2 | X_m = k_1\} = \frac{1}{2}$$

(ii) $P\{X_{3m+3} = 1 | X_{3m+2} = 1, X_{3m+1} = 1\} = 1$;

(iii) 上述的 $P\{X_n = k_2 | X_m = k_1\}$ 满足切普曼-柯尔莫哥洛夫方程，但 $\{X_n, n\geq 1\}$ 不是马尔可夫过程.

3. 两个状态的马尔可夫链. 设 $\{X_n, n\geq 1\}$ 是一个齐次两个状态 $\{0, 1\}$ 的马尔可夫链，它的转移概率的矩阵形式是：

$$P = \begin{pmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{pmatrix}$$

其中

$$p_{ij} = P\{X_{n+1} = j | X_n = i\}, \quad i, j = 0, 1$$

试证：(i) 两步转移概率矩阵是

$$P^{(2)} = P^2 = \begin{bmatrix} p_{00}^2 + p_{01}p_{10} & p_{01}(p_{00} + p_{11}) \\ p_{10}(p_{00} + p_{11}) & p_{1}^2 + p_{01}p_{10} \end{bmatrix}$$

(ii) 当 $|p_{00} + p_{11} - 1| < 1$ 时，试用数学归纳法证明 n 步转移概率矩

阵是

$$P^{(n)}=P^n=\frac{1}{2-p_{00}-p_{11}}\begin{bmatrix}1-p_{11} & 1-p_{00}\\ 1-p_{11} & 1-p_{00}\end{bmatrix}$$

$$+\frac{(p_{00}+p_{11}-1)^n}{2-p_{00}-p_{11}}\begin{bmatrix}1-p_{00} & -(1-p_{00})\\ -(1-p_{11}) & 1-p_{11}\end{bmatrix}$$

(iii) 由(ii)进一步证明

$$\lim_{n\to\infty}p_{00}^{(n)}=\lim_{n\to\infty}p_{10}^{(n)}=\frac{1-p_{11}}{2-p_{00}-p_{11}},$$

$$\lim_{n\to\infty}p_{01}^{(n)}=\lim_{n\to\infty}p_{11}^{(n)}=\frac{1-p_{00}}{2-p_{00}-p_{11}}.$$

(iv) 特别当 $p_{00}=p_{11}=p$, $p_{01}=p_{10}=q=1-p$ 时，有

$$P^{(n)}=\begin{bmatrix}\frac{1}{2}+\frac{1}{2}(p-q)^n & \frac{1}{2}-\frac{1}{2}(p-q)^n\\ \frac{1}{2}-\frac{1}{2}(p-q)^n & \frac{1}{2}+\frac{1}{2}(p-q)^n\end{bmatrix}$$

这时可以有如下的一个概率解释. 考虑一个数据通信系统，它由 n 个中继站组成. 从这一站向另一个站传送信号(0 或 1)时，在接收站以概率 p 正确接收，而以概率 $q=1-p$ 发生错误(即 0 变 1，或 1 变 0). 如用 X_0 表示初始站发出的数字，X_n 表示经 n 次传送后接收到的数字，$\{X_n\}$ 即构成上述两状态马尔可夫链.

(v) 在(iv)的条件下，试证

$$P\{X_0=1|X_n=1\}=\frac{\alpha+\alpha(p-q)^n}{1+(\alpha-\beta)(p-q)^n},$$

其中 $\alpha=P\{X_0=1\}$, $\beta=1-\alpha$. 并进一步说明上述概率的实际意义.

4. 设$\{X_n, n\geqslant 1\}$是一个齐次有限马尔可夫链，它的转移概率是(p_{ij})，如果还满足：

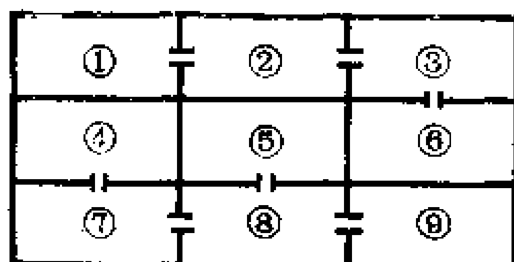
$$\sum_i p_{ij}=1,\quad 对一切\ j$$

则(p_{ij})称为**双重随机的**. 试证，如果$\{X_n, n\geqslant 1\}$还是非周期不可约的，则有

$$\lim_{n\to\infty}p_{ij}^{(n)}=\frac{1}{N},\quad 对一切\ i,j$$

其中N是状态的个数.

5. 将小白鼠放在如下的迷宫中，假定小白鼠在其中作随机的移动，即当



它处于某一格子中，而此格子又有 k 条路通入别的格子，则小白鼠以$\frac{1}{k}$的概率选择任一条路．如设小白鼠每次移动一个格子，并用 X_n 表示经 n 次移动后它所在的格子号码．试

(i) 说明$\{X_n, n\geqslant 1\}$构成一个齐次有限马尔可夫链；

(ii) 写出它的转移概率；

(iii) 分解它的状态空间$\{1, 2, \cdots, 7, 8, 9\}$.

6. 设$\{U_n, n\geqslant 1\}$为相互独立同分布的随机变量，其分布为

$$P\{U_n=1\}=p>0, \quad P\{U_n=0\}=q=1-p>0$$

(i) 定义$\{X_n, n\geqslant 2\}$如下：

$$X_n=\begin{cases}0, & \text{如 } U_n=0, U_{n-1}=0\\ 1, & \text{如 } U_n=0, U_{n-1}=1\\ 2, & \text{如 } U_n=1, U_{n-1}=0\\ 3, & \text{如 } U_n=1, U_{n-1}=1\end{cases}$$

试说明$\{X_n, n\geqslant 2\}$构成一个不可约非周期齐次马尔可夫链，并写出它的转移概率.

(ii) 定义$\{Y_n, n\geqslant 1\}$如下：

$$Y_n=\begin{cases}0, & \text{如 } U_n=0, U_{n-1}=0\\ 1, & \text{其它}\end{cases}$$

试说明$\{Y_n, n\geqslant 2\}$不是马尔可夫链.

顺便指出，若取定义在$\{0, 1, 2, 3\}$上的函数 f 为

$$f(x)=\begin{cases}0, & x=0\\ 1, & x=1, 2, 3\end{cases}$$

则 Y_n 可写为 $Y_n=f(X_n)$. 因此，一般的说，一个马尔可夫链的函数不再是一个马尔可夫链.

7. 设某金工车间有 M 台车床. 由于经常需要测量和调换刀具等原因，各车床总是时而停止，时而工作. 假定在时刻 t 时，一台车床正在工作，但在

时刻 $t+\Delta t$ 时停止工作的概率为 $\mu\Delta t+o(\Delta t)$，再假定在时刻 t 时，一台车床不工作，而在时刻 $t+\Delta t$ 时开始工作的概率为 $\lambda\Delta t+o(\Delta t)$，而且各车床的工作情况是相互独立的.

如用 $N(t)$ 表示时刻 t 正在工作的车床数，试

(i) 说明 $\{N(t), t\geqslant 0\}$ 是一个齐次马尔可夫过程；

(ii) 求出它的平稳分布 $\{p_i\}$；

(iii) 特别当 $M=10$，$\lambda=60$，$\mu=30$ 时，求出平稳态时有一半以上的车床在工作的概率.

8. 等待系统. 考虑一个服务站，其中有 M 个服务员. 顾客不断来到请求服务，假定在 Δt 时间中，来到一个顾客的概率为 $\lambda\Delta t+o(\Delta t)$，没有顾客来到的概率为 $1-\lambda\Delta t+o(\Delta t)$，来到两个或两个以上顾客的概率为 $o(\Delta t)$. 来到的顾客当有空着的服务员时，立即接受服务，否则排队等待，直到有一正在服务的顾客结束服务，服务员空出来后再去为排队的顾客服务. 假定每个在时刻 t 正在接受服务的顾客在时刻 $t+\Delta t$ 时结束服务的概率为

$$\mu\Delta t+o(\Delta t),$$

并假定各顾客的来到时刻，他们的服务时间等都是相互独立的.

如用 $N(t)$ 表示时刻 t 时正在排队的顾客个数(也就是队长)，试

(i) 说明 $N(t)$ 是一个马尔可夫过程；

(ii) 列出 $p_{ij}(t)=P\{N(t)=j\,|\,N(0)=i\}$ 所满足的方程；

(iii) 求出它的平稳分布.

9. 用上题的记号，进一步假定

(i) $M=1, \lambda=18, \mu=20$；

(ii) $M=1, \lambda=18, \mu=40$；

(iii) $M=1, \lambda=9, \mu=20$；

(iv) $M=2, \lambda=18, \mu=20$；

(v) $M=2, \lambda=18, \mu=10$；

试对上述五种情况分别求出平稳态时下列一些量：

(i) 一个顾客来到而无需等待的概率；

(ii) 一个顾客来到而需等待两个或两个以上顾客服务完后才接受服务的概率；

(iii) 平均队长.

10. 在第 8 题的假定下，试以平稳时的平均队长为指标，比较下列两个

系统：

$$M_1=2M_2,\quad \lambda_1=\lambda_2,\quad 2\mu_1=\mu_2$$

其中 M_1、λ_1，μ_1 和 M_2，λ_2，μ_2 分别为两个系统的相应的 M，λ，μ。

11. 用本章的爱尔兰公式，对下列五种情况分别计算：

(i) 用户的要求因线路被占满而被取消的概率；

(ii) 正在通话的线路平均数.

这五种情况是：(i) $s=10$, $\lambda=100$, $\mu=12$；

(ii) $s=10$, $\lambda=200$, $\mu=12$；

(iii) $s=10$, $\lambda=100$, $\mu=24$；

(iv) $s=5$, $\lambda=100$, $\mu=12$；

(v) $s=5$, $\lambda=100$, $\mu=24$.

12. 考虑一个细菌群体. 在一段时间内，假定它可以通过分裂等方法产生新的细菌，但不会死去. 假定在长为 Δt 的一段时间内，一个细菌分裂为两个，即产生新的一个的概率为 $\lambda\Delta t+o(\Delta t)$. 如用 $N(t)$表示时刻 t 时群体的大小，试(i) 说明 $N(t)$是一个生灭过程；

(ii) 证明：$\lambda_i=i\lambda$, $\mu_i=0$，并列出它的前进后退方程；

(iii) 验证：

$$p_{kj}(t)=\binom{j-1}{j-k}(e^{-\lambda t})^k(1-e^{-\lambda t})^{j-k},\quad j\geqslant k\geqslant 1$$

是上述方程的解，并计算 $E\{N(t+s)-N(s)\mid N(s)=m\}$.

注：由于 $\mu_i=0$，所以这过程也称**纯生过程**，又由于 $\lambda_i=i\cdot\lambda$，即 λ_i 与 i 成线性关系，所以进一步称为具有**线性生长率的纯生过程**. 相应地 $\lambda_i=0$ 的生灭过程称为纯灭过程.

13. (续上题)设 $N(0)\equiv1$，试求：

(i) $\mathrm{Var}\{N(t)\}$；

(ii) $\mathrm{cov}\{N(s), N(t)\}$；

其中 $\mathrm{Var}\{\cdot\}$，$\mathrm{cov}\{\cdot,\cdot\}$分别表示方差和协方差。

再设 $N(0)$为几何分布，即

$$P\{N(0)=k\}=p(1-p)^{k-1},\quad k=0,1,2,\cdots$$

其中 $0<p<1$. 试求

(iii) $E\{N(t)\}$；

(iv) $\mathrm{Var}\{N(t)\}$。

14. 首次超出时间与截尾过程.

设 $X(t)$ 是一齐次扩散过程, $X(0)=x_0$. a, b 是任意两个实数, 满足:

$$-\infty\leqslant a<x_0<b\leqslant\infty,$$

称下列随机变量

$$T\equiv T_{ab}(x_0)=\inf\{t:X(t)\leqslant a \text{ 或 } X(t)\geqslant b\}$$

为 $X(t)$ 首次超出区间 (a,b) 的时间. 特别当 $a=-\infty$ 或 $b=+\infty$ 时, 记为 $T_b(x_0)$ 和 $T_a(x_0)$. 这是实际问题中感兴趣的量. 求它的一个有用方法是引入所谓**截尾过程** $\overline{X}(t)$:

$$\overline{X}(t)=\begin{cases}X(t), & \text{如 } t<T\\ X(T), & \text{如 } t\geqslant T\end{cases}$$

即在 $X(t)$ 未超出区间 (a, b) 时, $\overline{X}(t)$ 与 $X(t)$ 一致, 而以后 $\overline{X}(t)$ 则恒等于 $X(T)$. 可以证明, 在一定条件下, $\overline{X}(t)$ 也是扩散过程.

进一步假设 $X(t)$ 是**对称过程**, 即其转移概率满足:

$$F(t,x_0,x)=1-F(t,-x_0,-x)$$

特别对 $a=0$, $b=\infty$, 即

$$T\equiv\inf\{t:X(t)\leqslant 0\}, \quad X(0)=x_0>0$$

试(i)证明

$$F(t,x_0,0)=\int_0^t G(d\tau,x_0,0)F(t-\tau,0,0)$$

其中 $G(t,x_0,0)=P\{T<t\}$ 是首次超出时间的分布函数.

(ii) 利用过程的对称性, 证明

$$F(t,0,0)=\frac{1}{2}.$$

(iii) 利用(i), (ii)证明:

$$G(t,x_0,0)=P\{T<t\}=2\cdot F(t,x_0,0)$$

(iv) 如 $\dfrac{\partial F(t,x_0,0)}{\partial t}$ 存在, 试证

$$g(t,x_0,0)=\frac{\partial G(t,x_0,0)}{\partial t}$$

存在, 且

$$g(t,x_0,0)=2\frac{\partial F(t,x_0,0)}{\partial t}$$

15. (续上题) 设 $X(t)$ 是维纳-爱因斯坦过程. 试说明:

(i) 它是对称过程；

(ii) $P\{\sup\limits_{0\leqslant\tau\leqslant t} X(\tau)\geqslant x\}=2P\{X(t)\geqslant x\}$；

(iii) $g(t, x_0, 0)=\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{x_0^2}{4Dt}\right)\frac{x_0}{\sqrt{2Dt^3}}$.

第三章　二阶矩过程和随机分析

本章中将继续研究第一章中提到的二阶矩过程 $X(t)$，特别我们将讨论 $X(t)$的连续性，导数和积分等．换句话说，我们要将普通分析的结果推广到二阶矩过程的场合，所以这部分内容可以称之为随机分析．

众所周知，连续性，导数和积分等概念的基础是极限的概念．在随机分析中也一样，为了确定二阶矩过程 $X(t)$的连续性，导数和积分，也必须先定义随机变量的极限．随机变量的极限有许多定义，本章中讨论的是所谓的均方极限，因此，更确切地说，本章中主要研究二阶矩过程的均方极限意义下的随机分析．

如第一章中所述，对于二阶矩过程 $X(t)$，不妨设其均值函数 $m(t)=E\{X(t)\}=0$，相应地

$$\Gamma(s,t)=E\{X(s)\overline{X(t)}\}$$

今后一般都这样假定的．

在§1中，我们介绍一些必要的预备知识．§2介绍随机分析．§3对正态过程的场合更进一步的分析．§4介绍伊藤随机积分．

§1. 预备知识

一、二阶矩随机变量全体所组成的空间 H

由于我们在这一章中只讨论二阶矩过程 $X(t)$，所以要反复处理各种具有二阶矩的随机变量 Y，Z，$X(t_1)$，$X(t_2)$，…．为方便起

见，今后我们称概率空间$(\Omega, \mathscr{F}, P)$上具有二阶矩的（复）随机变量为**二阶矩（随机）变量**，其全体记为H.

定理 1 设$X_1, X_2 \in H$, C_1, C_2为常数（可以是复的），则

$$C_1X_1+C_2X_2\in H$$

［证明］ 由许瓦兹不等式，有

$$(E\{|X_1X_2|\})^2 \leqslant E\{|X_1|^2\}\cdot E\{|X_2|^2\}<\infty$$

所以

$$E\{|C_1X_1+C_2X_2|^2\}\leqslant|C_1|^2E\{|X_1|^2\}$$
$$+2|C_1||C_2|E\{|X_1X_2|\}+|C_2|^2E\{|X_2|^2\}<\infty$$

系 H是一个线性空间.

我们今后将称两个随机变量X和Y是**等价的或相等的**，如果$P\{X\neq Y\}=0$，或$P\{X=Y\}=1$，并记作$X=Y$.

在上述定义下，根据许瓦兹不等式，对任意$X, Y\in H$，有：

$$|E\{X\overline{Y}\}|^2\leqslant E\{|X|^2\}E\{|Y|^2\}<\infty \tag{1}$$

因此可以引入

$$(X, Y)=E\{X\overline{Y}\}$$

且易证：

定理 2 (X, Y)是H中的内积，即如$X_1, X_2, X, Y\in H$，有

(i) $(Y, X)=\overline{(X, Y)}$;

(ii) $(CX, Y)=C(X, Y)$, $(X, CY)=\bar{C}(X, Y)$;

(iii) $(X_1+X_2, Y)=(X_1, Y)+(X_2, Y)$;

(iv) $(X, X)\geqslant 0$，当且仅当$X=0$时，$(X, X)=0$.

如对$X\in H$，引入记号

$$\|X\|=(X, X)^{\frac{1}{2}}$$

则有

引理 对于$X, Y\in H$，有

(i) $|E\{X\}|\leqslant E\{|X|\}\leqslant\|X\|$;

(ii) $|(X,Y)|=|E\{X\bar{Y}\}|\leqslant\|X\|\|Y\|$.

[证明] (ii)中取 $Y=1$ 即为(i), (ii)就是上面的(1)式。进而还有：

定理 3 $\|X\|$是H中的范数，即如 $X_1, X_2, X\in H$，有

(i) $\|X\|\geqslant 0$，当且仅当 $X=0$ 时，$\|X\|=0$；

(ii) $\|CX\|=|C|\|X\|$；

(iii) $\|X_1+X_2\|\leqslant\|X_1\|+\|X_2\|$.

[证明] (i), (ii)显然，现证(iii)，

$$
\begin{aligned}
\|X_1+X_2\|^2 &= (X_1+X_2, X_1+X_2)\\
&=(X_1,X_1)+(X_1,X_2)+(X_2,X_1)+(X_2,X_2)\\
&\leqslant\|X_1\|^2+|(X_1,X_2)|+|(X_2,X_1)|+\|X_2\|^2\\
&\leqslant\|X_1\|^2+\|X_1\|\|X_2\|+\|X_2\|\|X_1\|+\|X_2\|^2\\
&=(\|X_1\|+\|X_2\|)^2
\end{aligned}
$$

两边开方即得所证。

对 $X, Y\in H$，还可以引入

$$d(X,Y)=\|X-Y\|$$

于是有

定理 4 $d(X,Y)$是H中的距离，即对 $X,Y,Z\in H$，有

(i) $d(X,Y)\geqslant 0$，当且仅当 $X=Y$ 时，$d(X,Y)=0$；

(ii) $d(X,Y)=d(Y,X)$；

(iii) 三角不等式：$d(X,Z)\leqslant d(X,Y)+d(Y,Z)$.

有了上述的内积，范数，距离等概念，即可引入极限的概念。对 $X_n\in H, n\geqslant 1$ 和 $X\in H$，如有

$$\lim_{n\to\infty}d(X_n,X)=\lim_{n\to\infty}\|X_n-X\|=0$$

则称 X_n 均方收敛于 X，或 X 是 X_n 的均方极限，记作

$$\lim_{n\to\infty}x_n=X$$

仿普通分析，我们称 H 中的 $\{X_n, n\geqslant 1\}$ 为均方收敛的基本序列或柯西(Cauchy)序列，如有

$$\lim_{m,n\to\infty} d(X_m, X_n) = \lim_{m,n\to\infty} \|X_m - X_n\| = 0$$

容易证明

定理 5 如 $\lim\limits_{n\to\infty} X_n = X$，则 $\{X_n, n\geqslant 1\}$ 是基本序列。

反过来，我们还有：

定理 6 如 $\{X_n, n\geqslant 1\}$ 是基本序列，则存在 $X\in H$，使

$$\lim_{n\to\infty} X_n = X$$

这一定理也称为完备性定理，其证明要用到测度论知识，这儿从略。

综上所述，二阶矩随机变量全体 H 是一个完备的赋范线性空间，即巴拿赫空间，也是一个完备的有内积的线性空间，即希尔伯特空间。

现在如设 $\{X(t), t\in T\}$ 是一个二阶矩过程，则对每一 $t\in T$，$X(t)$ 是一个二阶矩随机变量，亦即是 H 中的一个元素，于是整个二阶矩过程 $\{X(t), t\in T\}$ 可以看作为希尔伯脱空间 H 中的一条曲线①。

二、均方极限的一些性质

除上述的定理 5 和定理 6 之外，还有：

定理 7 设 $X_n\in H$，$n\geqslant 1$，$X\in H$，且

$$\lim_{n\to\infty} X_n = X$$

则有

$$\lim_{n\to\infty} E\{X_n\} = E\{X\} = E\{\lim_{n\to\infty} X_n\}$$

---

① 本章和第四，五两章都将运用泛函分析中有关希尔伯脱空间的一些基本概念和结论，不熟悉这方面结果的读者可参看有关的教科书，例如"泛函分析"，南京大学数学天文学系编(1961)；"实变函数和泛函分析概要"，复旦大学数学系主编(1978)。

$$\lim_{n\to\infty} E\{|X_n|^2\} = E\{|X|^2\} = E\{|\lim_{n\to\infty} X_n|^2\}$$

［证明］ 由定理 2 和上段引理，有

$$\begin{aligned}|E\{X_n\} - E\{X\}| &= |E\{X_n - X\}| \\ &\leqslant E|X_n - X| \\ &\leqslant \|X_n - X\| \to 0\end{aligned}$$

最后一个不等式由许瓦兹不等式得出．另一式可由下式推出：

$$|\|X_n\| - \|X\|| \leqslant \|X_n - X\| \to 0$$

理定 8 设 $X_n, Y_n, X, Y \in H$，且

$$\lim_{m\to\infty} X_m = X, \qquad \lim_{n\to\infty} Y_n = Y$$

则有

$$\lim_{\substack{m\to\infty \\ n\to\infty}} (X_m, Y_n) = (X, Y)$$

或者

$$\lim_{\substack{m\to\infty \\ n\to\infty}} E\{X_m \overline{Y}_n\} = E\{X\overline{Y}\}$$

［证明］ 由定理 2 与上段引理，我们有：

$$\begin{aligned}&|(X_m, Y_n) - (X, Y)| \\ &= |(X, Y_n - Y) + (X_m - X, Y) + (X_m - X, Y_n - Y)| \\ &\leqslant \|X\|\|Y_n - Y\| + \|X_m - X\|\|Y\| + \|X_m - X\|\|Y_n - Y\|\end{aligned}$$

按定理条件 $\|X_m - X\| \to 0$，$\|Y_n - Y\| \to 0$，即得定理．

定理 9 设 $X_n, Y_n, X, Y \in H$，且

$$\lim_{n\to\infty} X_n = X, \qquad \lim_{n\to\infty} Y_n = Y$$

则对任意常数 a, b，有

$$\lim_{n\to\infty} (aX_n + bY_n) = aX + bY$$

［证明］ 由定理 3

$$\|aX_n + bY_n - (aX + bY)\| = \|a(X_n - X) + b(Y_n - Y)\|$$

$$\leqslant |a| \|X_n - X\| + |b| \|Y_n - Y\| \to 0$$

定理 10 均方极限是唯一的，即如设 $X_n, X, Y \in H$，且

$$\lim_{n\to\infty} X_n = X, \qquad \lim_{n\to\infty} X_n = Y$$

则有

$$X = Y$$

［证明］ 由定理 2，我们有

$$(X_n - Y, X_n - Y)$$
$$= (X_n, X_n) - (X_n, Y) - (Y, X_n) + (Y, Y)$$

两边取 $n\to\infty$，左边用 $X_n \to X$，右边用 $X_n \to Y$，由定理 9 和定理 8，得

$$(X - Y, X - Y)$$
$$= (Y, Y) - (Y, Y) - (Y, Y) + (Y, Y) = 0$$

再由定理 2 知 $X - Y = 0$，即 $X = Y$。

定理 11（均方收敛准则） 设 $X_n \in H$，$X \in H$，则 $\lim\limits_{n\to\infty} X_n = X$ 的充要条件是下列极限存在：

$$\lim_{\substack{m\to\infty \\ n\to\infty}} E\{X_m \overline{X}_n\} = \lim_{\substack{m\to\infty \\ n\to\infty}} (X_m, X_n) = C$$

此时，有

$$\lim_{\substack{m\to\infty \\ n\to\infty}} E\{X_m \overline{X}_n\} = E\{|X|^2\} = \|X\|^2$$

［证明］ 必要性由定理 9 立即推出，现证充分性．我们有

$$\|X_m - X_n\|^2 = (X_m - X_n, X_m - X_n)$$
$$= (X_m, X_m) - (X_m, X_n) - (X_n, X_m) + (X_n, X_n)$$

根据定理条件，在 $m, n \to \infty$ 时，右边趋于

$$C - C - C + C = 0$$

所以 $\{X_n\}$ 是基本序列，根据定理 6，存在 $X \in H$，使

$$\lim_{n\to\infty} X_n = X$$

此时由上述必要性,显然有

$$\lim_{\substack{m\to\infty\\ n\to\infty}} E\{X_m\bar{X}_n\}=\lim_{\substack{m\to\infty\\ n\to\infty}}(X_m,X_n)$$

$$=(X,X)=\|X\|^2$$

定理 12 设 X_n, $X\in H$, $\lim\limits_{n\to\infty}X_n=X$, $f(u)$是普通的函数, 使 $f(X_n),f(X)\in H$,且满足李不西兹(Lipschitz)条件:

$$|f(u)-f(v)|\leqslant M\cdot|u-v|$$

其中M为某一常数,则

$$\lim_{n\to\infty}f(X_n)=f(X)$$

[证明] 我们有

$$|f(X_n)-f(X)|^2\leqslant M^2|X_n-X|^2$$

于是

$$\|f(X_n)-f(X)\|^2=E|f(X_n)-f(X)|^2$$
$$\leqslant M^2E|X_n-X|^2=M^2\|X_n-X\|^2\to 0$$

系 1 如 $f(u)$和 $f'(u)$都存在且有界,以及 $\lim\limits_{n\to\infty}X_n=X$,则

$$\lim_{n\to\infty}f(X_n)=f(X)$$

系 2 设 $\lim\limits_{n\to\infty}X_n=X$, 则对任意有限的 t,

$$\lim_{n\to\infty}\exp(itX_n)=\exp(itX)$$

亦即若 X_n 均方收敛于 X, 则 X_n 的特征函数收敛于X的特征函数,从而 X_n 的分布函数收敛于X的分布函数。

下面我们来看一个简单的例子.

设 Y_n, $n\geqslant 1$ 是独立同分布的二阶矩随机变量,其均值记为 m

$$E\{X\}=m$$

则由定理 1 可知,

$$X_n=\frac{1}{n}\sum_{k=1}^{n}Y_k$$

也是二阶矩随机变量，我们要证明它均方收敛于均值 m. 事实上有

$$\|X_n-m\|^2=\left\|\frac{1}{n}\sum_{k=1}^{n}Y_k-m\right\|^2$$

$$=\left\|\frac{1}{n}\sum_{k=1}^{n}(Y_k-m)\right\|^2=\frac{1}{n^2}\left\|\sum_{k=1}^{n}(Y_k-m)\right\|^2$$

$$=\frac{1}{n^2}E\left\{\sum_{k=1}^{n}(Y_k-m)\cdot\sum_{l=1}^{n}(Y_l-m)\right\}$$

利用 Y_n 的独立性，诸交叉项

$$E\{(Y_k-m)(Y_l-m)\}$$

$$=E\{Y_k-m\}E\{Y_l-m\}=0$$

所以有

$$\|X_n-m\|^2=\frac{1}{n^2}\sum_{k=1}^{n}E\{|Y_k-m|^2\}$$

$$=\frac{1}{n}E\{|Y_1-m|^2\}\to 0$$

上述结论也可写成:

定理 13　均方极限意义下的大数定律　设 Y_n 是独立同分布的随机变量，且其二阶矩存在，则

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}Y_k=E\{Y\}$$

§2. 随机分析

本节中研究二阶矩过程 $\{X(t),\ t\in T\}$ 的微积分，为确定起见，假定 $T=[a,b]$ 是直线上某一区间(可以是无穷)，且恒假定其均值函数 $m(t)=0$，其协方差函数为:

$$\Gamma(s,t)=E\{X(s)\overline{X(t)}\}$$

一、均方连续性

二阶矩过程$\{X(t), t\in T\}$称为在 t_0 **均方连续的**，如果有

$$\lim_{t\to t_0} X(t)=X(t_0)$$

亦即

$$\lim_{t\to t_0} \|X(t)-X(t_0)\|=0$$

如果 $X(t)$对 T 中的一切 t 都连续，则称$\{X(t), t\in T\}$**在 T 上均方连续**.

定理 1 （均方连续准则）$\{X(t), t\in T\}$在 t_0 处均方连续的充要条件是：$\Gamma(s,t)$在(t_0,t_0)处连续.

［证明］ 由定理 1.11 可知，

$$\lim_{t\to t_0} X(t)=X(t_0)$$

的充要条件是

$$\lim_{\substack{s\to t_0\\ t\to t_0}} E\{X(s)\overline{X(t)}\}=E\{X(t_0)\overline{X(t_0)}\}$$

这就是

$$\lim_{\substack{s\to t_0\\ t\to t_0}} \Gamma(s,t)=\Gamma(t_0,t_0)$$

系 1 $\{X(t), t\in T\}$在 T 上均方连续的充要条件是：$\Gamma(s,t)$在$\{(t,t), t\in T\}$上二元连续.

系 2 如果协方差函数 $\Gamma(s,t)$在$\{(t,t), t\in T\}$上连续，则它在 $T\times T$ 上连续，即对协方差函数 $\Gamma(s,t)$而言，它在整个区域 $T\times T$ 上连续与它在 $T\times T$ 的对角线上连续是等价的.

［证明］ 由系的条件和定理可知，

$$\lim_{s\to s_0} X(s)=X(s_0)$$

$$\lim_{t\to t_0} X(t)=X(t_0)$$

于是应用定理 1.8 有

$$\lim_{\substack{s\to s_0\\ t\to t_0}} E\{X(s)\overline{X(t)}\}=E\{X(s_0)\overline{X(t_0)}\}$$

也就是

$$\lim_{\substack{s \to s_0 \\ t \to t_0}} \Gamma(s, t) = \Gamma(s_0, t_0)$$

二、均方导数

二阶矩过程$\{X(t), t \in T\}$称为在$t \in T$**是均方可微的**，如果下列极限存在

$$\lim_{h \to 0} \frac{X(t+h) - X(t)}{h}$$

此极限记作$\dot{X}(t)$或$X'(t)$或$\frac{dX(t)}{dt}$，称为$X(t)$在t处的**均方导数**，或均方微商.

如果$\{X(t), t \in T\}$在T上每一点都均方可微的，则称$\{X(t), t \in T\}$**在T上均方可微**.

如果$\{X'(t), t \in T\}$在t处均方可微，则称$\{X(t), t \in T\}$在t处**二次均方可微**. $X'(t)$的均方导数记为$X''(t)$，称为$X(t)$的**二阶均方导数**. 类似地，可以定义更高阶的导数.

为了叙述均方可微的准则，我们还需要下列**广义二阶导数**的概念. 一个普通的二元函数$f(s, t)$，称为在(s, t)处**广义二次可微**，如果下列极限存在

$$\lim_{\substack{h \to 0 \\ h' \to 0}} \frac{f(s+h, t+h') - f(s+h, t) - f(s, t+h') + f(s, t)}{h \cdot h'}$$

这一极限就称为$f(s, t)$在(s, t)处的**广义二阶导数**或**广义二次微商**.

我们指出，只要$f(s, t)$关于s和t的一阶偏导数存在，二阶混合偏导数$\frac{\partial^2}{\partial s \partial t} f(s, t)$存在且连续，则$f(s, t)$就是广义二次可微的. 但如没有上述的连续条件，则一般是不对的，一个简单的例子如下. 在$[0, 1] \times [0, 1]$上定义一个$f(s, t)$，它只取0或1两个

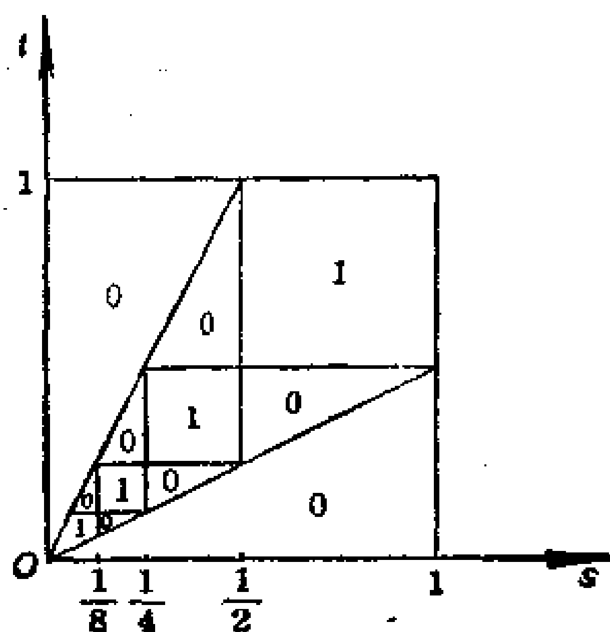


图 1

值，而取 0 和 1 值的区域如图 1 所示．考察(0, 0)这一点，显然有

$$\frac{\partial}{\partial s}f(0,0)=\frac{\partial}{\partial t}f(0,0)$$

$$=\frac{\partial^2}{\partial s\partial t}f(0,0)=\frac{\partial^2}{\partial t\partial s}f(0,0)$$

$$=0$$

但特别当 $h=h'\to 0$ 时，

$$\lim_{h\to 0}\frac{f(h,h)-f(h,0)-f(0,h)+f(0,0)}{h^2}$$

$$=\lim_{h\to 0}\frac{1}{h^2}=\infty$$

因此不是广义二次可微的．

现在可以叙述下面的定理：

定理 2（均方可微准则） $X(t)$，$t\in T$ 在 t 处均方可微的充要条件是：$\Gamma(s,t)$在(t,t)处广义二次可微．

［证明］ 由定理 1.11 知，$X(t)$在 t 处均方可微，即

$$\lim_{h\to 0}\frac{X(t+h)-X(t)}{h}$$

存在的充要条件是

$$\lim_{\substack{h\to 0\\ h'\to 0}} E\left\{\left(\frac{X(t+h)-X(t)}{h}\right)\overline{\left(\frac{X(t+h')-X(t)}{h'}\right)}\right\}$$

存在，将上式展开，也就是要求

$$\lim_{\substack{h\to 0\\ h'\to 0}} \frac{\Gamma(t+h, t+h')-\Gamma(t, t+h')-\Gamma(t+h, t)+\Gamma(t, t)}{h\cdot h'}$$

存在，亦即要求 $\Gamma(s, t)$在(t, t)处广义二次可微.

系 1 $\{X(t), t\in T\}$在T上均方可微的充要条件是$\Gamma(s, t)$在一切$\{(t, t), t\in T\}$上广义二次可微.

系 2 如果$\Gamma(s, t)$在一切$\{(t, t), t\in T\}$上广义二次可微，则$\frac{\partial}{\partial s}\Gamma(s, t)$, $\frac{\partial}{\partial t}\Gamma(s, t)$, $\frac{\partial^2}{\partial s\partial t}\Gamma(s, t)$, $\frac{\partial^2}{\partial t\partial s}\Gamma(s, t)$ 在 $T\times T$ 上都存在，且有

$$E\{X'(s)\overline{X(t)}\}=\frac{\partial}{\partial s}\Gamma(s, t)$$

$$E\{X(s)\overline{X'(t)}\}=\frac{\partial}{\partial t}\Gamma(s, t)$$

$$E\{X'(s)\overline{X'(t)}\}=\frac{\partial^2}{\partial s\partial t}\Gamma(s, t)$$

$$=\frac{\partial^2}{\partial t\partial s}\Gamma(s, t)$$

【证明】 由系 1, $X'(t)$, $t\in T$ 存在. 于是根据§1定理8，有

$$E\{X'(s)\overline{X(t)}\}=\lim_{h\to 0} E\left\{\left(\frac{X(s+h)-X(s)}{h}\right)\overline{X(t)}\right\}$$

$$=\lim_{h\to 0}\frac{\Gamma(s+h, t)-\Gamma(s, t)}{h}=\frac{\partial}{\partial s}\Gamma(s, t)$$

其余部分可类似地证明.

下面再给出均方导数的性质，其证明与普通分析中是类似的,

故从略.

定理 3 设$X(t)$在t处均方可微, 则$X(t)$在t处均方连续.

定理 4 均方导数是唯一的, 即如$X'(t)=X$, $X'(t)=Y$, 则$X=Y$.

定理 5 设$X(t)$和$Y(t)$均方可微, a, b为常数, 则$aX(t)+bY(t)$也均方可微, 且有:

$$(aX(t)+bY(t))'=aX'(t)+bY'(t)$$

定理 6 设 $X(t)$ 均方可微, $f(t)$ 是普通的可微函数, 则$f(t)X(t)$均方可微, 且有

$$\frac{d}{dt}[f(t)X(t)]=\frac{df(t)}{dt}X(t)+f(t)\frac{dX(t)}{dt}$$

三、均方积分

设$X(t)$, $t\in T$是二阶矩过程, $f(t)$, $t\in T$为一个普通的函数, 考虑$T=[a,b]$的一组分点:

$$a=t_0<t_1<t_2<\cdots<t_n=b$$

$$\Delta_n=\max_{1\leqslant k\leqslant n} t_k-t_{k-1}$$

和下列和式

$$Y_n=\sum_{k=1}^{n}f(u_k)X(u_k)(t_k-t_{k-1})$$

其中$t_{k-1}\leqslant u_k\leqslant t_k$, $1\leqslant k\leqslant n$. 由定理 1.1 可知$Y_n\in H$.

如果当$\Delta_n\to 0$时, Y_n均方收敛, 则称$f(t)X(t)$在$[a,b]$上(黎曼)均方可积, 其极限记为

$$\int_a^b f(t)X(t)dt$$

称为$f(t)X(t)$在$[a,b]$上的(黎曼)均方积分.

仿普通分析, 如果

$$\lim_{b\to\infty}\int_a^b f(t)X(t)dt$$

存在，则记为

$$\int_a^{\infty} f(t)X(t)dt = \lim_{b\to\infty}\int_a^b f(t)X(t)dt$$

定理 7（**均方可积准则**）　$f(t)X(t)$在$[a, b]$上均方可积的充分条件是下列普通的二重积分存在：

$$\int_a^b\int_a^b f(s)\overline{f(t)}\Gamma(s, t)dsdt$$

〔证明〕　由§1定理11知，$f(t)X(t)$在$[a,b]$均方可积，即

$$\lim_{\Delta_n\to 0}\sum_{k=1}^{n} f(u_k)X(u_k)(t_k-t_{k-1})$$

存在的充分条件是下列极限存在，

$$\lim_{\Delta_n\to 0} E\left\{\left[\sum_{k=1}^{n} f(u_k)X(u_k)(t_k-t_{k-1})\right]\overline{\left[\sum_{l=1}^{n} f(u_l)X(u_l)(t_l-t_{l-1})\right]}\right\}$$

$$=\lim_{\Delta_n\to 0}\sum_{k=1}^{n}\sum_{l=1}^{n} f(u_k)\overline{f(u_l)}E\{X(u_k)\overline{X(u_l)}\}(t_k-t_{k-1})(t_l-t_{l-1})$$

$$=\lim_{\Delta_n\to 0}\sum_{k=1}^{n}\sum_{l=1}^{n} f(u_k)\overline{f(u_l)}\Gamma(u_k,u_l)(t_k-t_{k-1})(t_l-t_{l-1})$$

亦即

$$\int_a^b\int_a^b f(s)\overline{f(t)}\Gamma(s, t)dsdt$$

存在。

系　$\int_a^{\infty} f(t)X(t)dt$存在的充分条件是下列二重积分存在：

$$\int_a^{\infty}\int_a^{\infty} f(s)\overline{f(t)}\Gamma(s, t)dsdt$$

定理 8　设$f(t)X(t)$在$[a,b]$上均方可积，则有

(i) $E\left\{\int_a^b f(t)X(t)dt\right\}=\int_a^b f(t)E\{X(t)\}dt=0$;

(ii) $E\left\{\left(\int_a^b f(s)X(s)\,ds\right)\left(\overline{\int_a^b f(t)X(t)\,dt}\right)\right\}$

$$=\int_a^b\int_a^b f(s)\overline{f(t)}\Gamma(s,t)dsdt.$$

[证明] 用定理1.8即可.

系 设$X(t)$, $t\in T=[a,b]$均方连续,则对一切$t\in T$,有

$$\left\|\int_a^t X(s)ds\right\|^2\leqslant(t-a)\int_a^t\|X(s)\|^2ds$$

$$\leqslant(b-a)\int_a^t\|X(s)\|^2ds$$

[证明] 第二个不等式是显然的,只要证第一个。根据(ii),有:

$$\left\|\int_a^t X(s)ds\right\|^2=\int_a^t\int_a^t\Gamma(s,r)dsdr$$

$$=\int_a^t\int_a^t E\{X(s)\overline{X(r)}\}dsdr$$

$$\leqslant\int_a^t\left\|X(s)\right\|ds\cdot\int_a^t\|X(r)\|ds$$

$$=\left(\int_a^t\left\|X(s)\right\|ds\right)^2$$

$$\leqslant\left(\int_a^t 1^2ds\right)\left(\int_a^t\left\|X(s)\right\|^2ds\right)$$

$$=(t-a)\int_a^t\|X(s)\|^2ds$$

下面再给出均方积分的一些性质.

定理9 设$X(t)$在$[a,b]$上均方连续,则$X(t)$在$[a,b]$上均方可积.

[证明] 由定理1及其系可知,$X(t)$在$T=[a,b]$上均方连续可推出$\Gamma(s,t)$在$T\times T$上连续,于是$\Gamma(s,t)$在$T\times T$上可积,即

$$\int_a^b \int_a^b \Gamma(s, t) ds dt$$

存在. 再由定理 7 即得定理.

定理 10 均方积分是唯一的.

[证明] 由定理 1.10 直接推出.

定理 11 设 $X(t)$在$[a, b]$上均方连续, 则

$$\left\| \int_a^b X(t) dt \right\| \leqslant \int_a^b \| X(t) \| dt$$

[证明] 由定理 9 知, 上面左边的积分存在, 再由定理 1.1 及其系, 知$\|X(t)\| = \sqrt{\Gamma(t, t)}$连续, 于是上面右边的积分存在.

因为

$$\left\| \int_a^b X(t) dt \right\| = \left\| \lim_{\Delta_n \to 0} \sum_k X(u_k)(t_k - t_{k-1}) \right\|$$
$$= \lim_{\Delta_n \to 0} \left\| \sum_k X(u_k)(t_k - t_{k-1}) \right\|$$

另一方面, 有

$$\left\| \sum_k X(u_k)(t_k - t_{k-1}) \right\| \leqslant \sum_k \| X(u_k) \| (t_k - t_{k-1})$$

$$\longrightarrow \int_a^b \| X(t) \| dt$$

定理于是得证.

系 如$\lim\limits_{n \to \infty} X_n(t) = X(t)$关于 t 一致, 则

$$\lim_{n \to \infty} \int_a^b X_n(t) dt = \int_a^b X(t) dt.$$

定理 12 设 $X(t)$, $Y(t)$在$[a, c]$上均方可积, α, β 为常数, 则有

(i) $\int_a^c [\alpha X(t) + \beta Y(t)] dt = \alpha \cdot \int_a^c X(t) dt + \beta \int_a^c Y(t) dt$;

(ii) $\int_a^c X(t) dt = \int_a^b X(t) dt + \int_b^c X(t) dt$, $a \leqslant b \leqslant c$.

[证明]　与普通分析中的证明相仿，从略。

定理 13　设 $X(t)$在$[a,b]$上均方连续，则

$$Y(t)=\int_a^t X(s)ds,\quad a\leqslant t\leqslant b$$

在$[a,b]$上均方连续，均方可微，且有

$$Y'(t)=X(t)$$

[证明]　均方连续是显然的，现在来证均方可微. 由上一定理，有：

$$\left\|\frac{Y(t+h)-Y(t)}{h}-X(t)\right\|$$

$$=\left\|\frac{1}{h}\int_t^{t+h} X(s)ds-X(t)\right\|$$

$$=\left\|\frac{1}{h}\int_t^{t+h} (X(s)-X(t))ds\right\|$$

$$\leqslant\frac{1}{h}\int_t^{t+h}\|X(s)-X(t)\|ds$$

$$\leqslant\max_{|t-s|\leqslant h}\|X(s)-X(t)\|\to 0,\quad (h\to 0)$$

系　设 $X(t)$均方可微，且 $X'(t)$均方连续，则有

$$X(b)-X(a)=\int_a^b X'(t)dt$$

上面介绍了(黎曼)均方积分的定义和基本性质，这些结果都可以推广到一个二阶矩过程 $X(t)$ 关于有界变差函数 $g(t)$ 的 (黎曼-司蒂吉斯)均方积分. 例如要定义积分$\int_a^b X(t)dg(t)$，可考虑$[a,b]$的一组分点：

$$a=t_0<t_1<\cdots<t_n=b$$

$$\Delta_n=\max_{1\leqslant k\leqslant n}(t_k-t_{k-1})$$

并作出和式

$$\sum_{k=1}^{n} X(u_k)[g(t_k)-g(t_{k-1})],\quad t_{k-1}\leqslant u_k\leqslant t_k$$

若当 $\Delta_n \to 0$ 时上述和式的均方极限存在，则称这一极限为 $X(t)$ 关于 $f(t)$ 在 $[a, b]$ 上的**（黎曼-司蒂吉斯）均方积分**，并记为 $\int_a^b X(t)df(t)$. 这时，定理 7——定理 12 除了定理 8 的系之外都可以推广到这一情形，只需将其中积分的微分元 dt, ds 分别改为 $dg(t), dg(s)$ 即可.

四、关于正交增量过程的积分

上面我们介绍了随机过程的均方积分，这里我们将介绍普通函数 $f(t)$ 关于正交增量过程 $X(t)$ 的积分，这是在平稳过程谱分解中十分有用的一种积分.

若 $T=[a, b]$ 为有限或无限区间，$X_T=\{X(t),\ t\in[a, b]\}$ 为零均值的正交增量过程，即

$$E\{X(t)\}=0$$

且对任一 $a\leqslant t_1<t_2\leqslant t_3<t_4\leqslant b$ 有

$$E\{[X(t_2)-X(t_1)]\overline{[X(t_4)-X(t_3)]}\}=0$$

在此我们附加条件

$$X(a)=0 \tag{1}$$

（若 $a=-\infty$ 时，则取为 $\lim\limits_{t\to-\infty} E|X(t)|^2=0$.）则由 § 1.3 的讨论可知，$X(t)$ 的协方差函数 $\Gamma(s, t)$ 可表为

$$\Gamma(s, t)=F(\min(s, t)) \tag{2}$$

且上述 $F(s)$ 是单调不减的，$F(a)=0$. 为了建立普通函数 $f(t)$ 关于 X_T 的积分，我们还需假定 X_T 是（均方）左连续的，即对任一 $t\in(a, b]$ 有

$$\lim_{s\to t-0} E|X(t)-X(s)|^2=0 \tag{3}$$

由(2)，当 $t>s$ 时，

$$E|X(t)-X(s)|^2=F(t)-F(s)$$

所以(3)与 $F(t)$的左连续性等价.

若以 $L_2(dF)$表示$[a, b]$上关于F平方可积的波雷尔可测函数全体, 即

$$L_2(dF)=\left\{f(t):\int_a^b|f(t)|^2dF(t)<\infty\right\}$$

对 $f\in L_2(dF)$,我们来考虑它关于左连续正交增量过程 X_T 的积分.

(i) 若 $a\leqslant c<d\leqslant b$, 对 $f(t)=\chi_{[c,d)}(t)\in L_2(dF)$, 我们规定

$$\int_a^b\chi_{[c,d)}(t)dX(t)=X(d)-X(c) \tag{4}$$

则不难验证

$$E\left\{\int_a^b \chi_{[c,d)}(t)dX(t)\cdot\overline{\int_a^b \chi_{[e,f)}(t)dX(t)}\right\}$$
$$=\int_a^b\chi_{[c,d)}(t)\chi_{[e,f)}(t)dF(t)$$

(ii) 若 $f(t)=\sum_{i=1}^n k_i\chi_{[c_i,d_i)}(t)\in L_2(dF)$, 我们规定

$$\int_a^b f(t)dX(t)=\sum_{i=1}^n k_i[X(d_i)-X(c_i)]$$

可以说明这样规定是合理的, 即若

$$f(t)=\sum_{i=1}^n k_i\chi_{[c_i,d_i)}=\sum_{j=1}^{n'}k'_j\chi_{[c'_j,d'_j)}$$

则

$$\sum_{i=1}^n k_i[X(d_i)-X(c_i)]=\sum_{j=1}^{n'}k'_j[X(d'_j)-X(c'_j)]$$

同时从这一规定出发, 容易推出当$f_1(t)=\sum_{i=1}^n k_i\chi_{[c_i,d_i)}(t)$, $f_2(t)=\sum_{j=1}^{n'}l_j\chi_{[c'_j,d'_j)}(t)$都是区间示性函数的线性组合时,

$$\int_a^b [\alpha f_1(t)+\beta f_2(t)]dX(t)$$
$$=\alpha\int_a^b f_1(t)dX(t)+\beta\int_a^b f_2(t)dX(t) \tag{5}$$

又因为

$$E\left\{\int_a^b f_1(t)\,dX(t)\cdot\overline{\int_a^b f_2(t)dX(t)}\right\}$$
$$=E\left\{\left[\sum_{i=1}^{n}k_i(X(d_i)-X(c_i))\right]\overline{\left[\sum_{j=1}^{n'}l_j(X(d_j')-X(c_j'))\right]}\right\}$$
$$=\sum_{i=1}^{n}\sum_{j=1}^{n'}k_i\overline{l_j}E\{[X(d_i)-X(c_i)]\overline{[X(d_j')-X(c_j')]}\}$$
$$=\sum_{i=1}^{n}\sum_{j=1}^{n'}k_i\overline{l_j}\int_a^b \chi_{[c_i,d_i)}(t)\chi_{[c_j',d_j')}(t)dF(t)$$
$$=\int_a^b f_1(t)\overline{f_2(t)}dF(t)$$

所以

$$E\left\{\int_a^b f_1(t)dX(t)\cdot\overline{\int_a^b f_2(t)dX(t)}\right\}$$
$$=\int_a^b f_1(t)\overline{f_2(t)}dF(t) \tag{6}$$
$$E\left|\int_a^b f_1(t)dX(t)\right|^2=\int_a^b |f_1(t)|^2dF(t) \tag{7}$$

(iii) 最后，对任一 $f(t)\in L_2(dF)$，由 $L_2(dF)$ 的构造可知必存在一列 $\{f_n\}$ 满足

$$\lim_{n\to\infty}\int_a^b |f(t)-f_n(t)|^2dF(t)=0$$

而每个 $f_n(t)$ 都是 (ii) 中提到的区间示性函数的线性组合，因而 $\int_a^b f_n(t)dX(t)$ 有确定的含义，而由(5)，(7)，

$$E\left|\int_a^b f_n(t)dX(t)-\int_a^b f_m(t)dX(t)\right|^2$$

$$= E\left|\int_a^b (f_n(t)-f_m(t))dX(t)\right|^2$$

$$= \int_a^b |f_n(t)-f_m(t)|^2 dF(t) \to 0 \qquad (m, n\to\infty)$$

所以利用H的完备性，必有二阶矩变量Y满足

$$\lim_{n\to\infty} E\left|\int_a^b f_n(t)dX(t) - Y\right|^2 = 0$$

这时我们就规定

$$Y = \int_a^b f(t)dX(t)$$

仍可以说明这样规定是合理的，即收敛于f的$\{f_n\}$取法不同并不影响Y。因而$Y=\int_a^b f(t)dX(t)$有确定的含义。而且利用定理1.8和1.9立即推出(5)——(7)对$f_1, f_2\in L_2(dF)$也是成立的。

上述建立积分的过程可归结为下列定理：

定理 14 若$X_T=\{X(t),\ t\in[a, b]\}$为零均值的左连续正交增量过程，且满足(1)及(2)，则对任一$f\in L_2(dF)$，可唯一地建立f关于X_T的均方随机积分$\int_a^b f(t)dX(t)$，使它满足(4)——(7)，且这时必有

$$E\left\{\int_a^b f(t)dX(t)\right\} = 0 \tag{8}$$

〔证明〕 积分存在唯一性前面已加证明。当$f(t)=\chi_{[c,d)}(t)$时，(8)可由X_T的均值为零推出，而对一般的f，(8)可由均值运算的线性性及定理1.7推出。

关于正交增量过程随机积分常用的一些性质大都可以从它的定义和上述(4)——(7)推出，我们仅指出以后要用到的两个事实。

定理 15 若$X_T=\{X(t), t\in(-\infty, \infty)\}$为满足定理14条件的正交增量过程，$f(t)\in L_2(dF)$为$(-\infty, \infty)$上连续函数，则

$$\int_{-\infty}^{\infty} f(t)dX(t)=\lim_{N\to\infty}\lim_{\max|t_j^{(N)}-t_{j-1}^{(N)}|\to 0}\sum_{j=1}^{k_N} f(\bar{t}_j^{(N)})$$

$$[X(t_j^{(N)})-X(t_{j-1}^{(N)})]$$

其中$\{t_j^{(N)}\}$为$[-N, N]$上的一组分点，满足

$$-N=t_0^{(N)}<t_1^{(N)}<\cdots<t_{k_N}^{(N)}=N$$

$$t_{j-1}^{(N)}\leqslant \bar{t}_j^{(N)}\leqslant t_j^{(N)}$$

［证明］ 若记$f_N(t)=\sum_{j=1}^{k_N} f(\bar{t}_j)\chi_{[t_{j-1}^{(N)}, t_j^{(N)})}(t)$，则

$$E\left|\int_{-\infty}^{\infty} f(t)dX(t)-\sum_{j=1}^{k_N} f(\bar{t}_j^{(N)})[X(t_j^{(N)})-X(t_{j-1}^{(N)})]\right|^2$$

$$=E\left|\int_{-\infty}^{\infty} f(t)dX(t)-\int_{-N}^{N} f_N(t)dX(t)\right|^2$$

$$\leqslant c\left\{E\left|\int_{-\infty}^{-N} f(t)dX(t)\right|^2+E\left|\int_N^{\infty} f(t)dX(t)\right|^2\right.$$

$$\left.+E\left|\int_{-N}^{N}(f(t)-f_N(t))dX(t)\right|^2\right\}$$

$$\leqslant c\left(\int_{-\infty}^{-N}|f|^2dF+\int_N^{\infty}|f|^2dF+\int_{-N}^{N}|f(t)-f_N(t)|^2dF(t)\right)$$

所以利用$f\in L_2(dF)$，可选取N使上式前两项之和小于ε，又对此固定的N，利用$f(t)$在$[-N, N]$上一致连续性，只要$\max\limits_{1\leqslant j\leqslant k_N}|t_j^{(N)}-t_{j-1}^{(N)}|$足够小，就可使第三项足够小，所以定理成立。

定理 16 若$X_T=\{X(t), t\in[a, b]\}$为满足定理 14 条件的正交增量过程，

(i) 若 $f\in L_2(dF)$，又$Y(t)=\int_a^t f(s)dX(s)$，则$Y_T=\{Y(t), t\in[a, b]\}$也是正交增量过程，且

$$E|Y(t)|^2=\int_a^t |f(s)|^2dF(s)\triangleq G(t)$$

又当$g\in L_2(dG)$时，

$$\int_a^b g(s)dY(s) = \int_a^b g(s)f(s)dX(s)$$

(ii) 若 $F(t)$是$[a, b]$上有界函数, $f(s, t)$是$[c, d]\times[a, b]$上二元连续函数, 且对每个 $s, f(s, \cdot)\in L_2(dF)$. 令

$$Y(s) = \int_a^b f(s, t)dX(t)$$

则 $Y_T = \{Y(s), s\in[c, d]\}$为均方连续二阶矩过程, 且当 c, d 为有限数时有

$$\int_c^d Y(s)ds = \int_a^b \left(\int_c^d f(s, t)ds\right)dX(t)$$

由于上述定理的证明可以用前面的方法获得, 所以留给读者作为练习.

上面我们是在(1)之下建立了关于正交增量过程的积分. 若借助于测度论中的测度扩张定理, 那么并不需要附加(1)也可以获得完全类似的结果①.

这一段和上一段定义了两种不同的均方积分. 一个自然的问题是当被积函数 $f(t)$也是随机过程时, 是否可以定义它关于另一个随机过程的积分以及是否有比较便于运算的性质, 这些问题将在 § 4 中再来讨论.

最后要说明一点, 我们在这一节定义了随机过程的均方连续和均方微分、积分, 在此极限都是取为均方收敛的意义, 若从随机过程的样本函数来看, 均方意义下的连续、可微或可积不能保证样本函数有相应的性质.

五、随机微分方程

有了随机微分和随机积分, 自然可以进一步讨论随机微分方程的问题. 这方面的全面讨论超出了本书的范围, 所以本段中只

① 参见 I. I. Gihman, A. V. Skorohod, The Theory of Stochastic Processes I, (1974).

讨论一些最简单的情形.

最简单的微分方程是

$$\begin{cases} X'(t)=Y(t), \qquad t\in T=[a, b] \\ X(t_0)=X_0 \end{cases}$$

其中 $Y(t)$是二阶矩过程, X_0 是二阶矩随机变量. 它的解由定理 13 及其系可知为:

$$X(t)=X_0+\int_{t_0}^{t} Y(s)ds$$

这个解也是唯一的. 由前面的讨论可知, 当$\{Y(t), t\in T\}$和 X_0 独立时, 有

$$E\{X(t)\}=E\{X_0\}+E\left\{\int_{t_0}^{t} Y(s)ds\right\}=E\{X_0\}$$

$$E\{X(s)\overline{X(t)}\}=E\{X_0X_0\}+E\{X_0\}\overline{E\left\{\int_{t_0}^{t} Y(s)ds\right\}}$$

$$+E\left\{\int_{t_0}^{s} Y(s)\,ds\right\}\overline{E\{X_0\}}+E\left\{\int_{t_0}^{s} Y(s)\,ds\overline{\int_{t_0}^{t} Y(s)ds}\right\}$$

$$=E\{|X_0|^2\}+\int_{t_0}^{s}\int_{t_0}^{t}\Gamma_Y(u, v)dudv$$

即$\{X(t), t\in T\}$的二阶矩完全由 X_0 和$\{Y(t), t\in T\}$ 的二阶矩所决定.

更复杂一些的一阶线性随机微分方程是

$$\begin{cases} X'(t)=a(t)X(t)+Y(t) \\ X(t_0)=X_0 \end{cases} \tag{1}$$

其中 $a(t)$是普通的函数, $Y(t)$是二阶矩过程, X_0是二阶矩随机变量.

为了讨论它的解, 我们先讨论下列相应的普通微分方程

$$\begin{cases} x'(t)=a(t)x(t)+y(t) \\ x(t_0)=x_0 \end{cases} \tag{2}$$

如果 $a(t), y(t)$是连续的, 则可以直接验证它的解是

$$x(t)=x_0\exp\left\{\int_{t_0}^{t}a(\tau)d\tau\right\}$$
$$+\int_{t_0}^{t}\exp\left\{\int_{s}^{t}a(\tau)d\tau\right\}y(s)ds$$

特别当 $a(t)=a=$ 常数时,

$$x(t)=x_0e^{a(t-t_0)}+\int_{t_0}^{t}e^{a(t-s)}y(s)ds$$

相应地,我们有下列结果.

定理 17 设 $a(t)$ 是普通的连续函数, $Y(t)$ 是均方连续的二阶矩过程, X_0 是二阶矩随机变量,则一阶线性随机微分方程(1)有解,且其解可表示为

$$X(t)=X_0\exp\left\{\int_{t_0}^{t}a(\tau)d\tau\right\}$$
$$+\int_{t_0}^{t}\exp\left\{\int_{s}^{t}a(\tau)d\tau\right\}Y(s)ds$$

其中关于 $Y(s)$ 的积分为均方积分.

[证明] 直接验证即可. $X(t_0)=X_0$. 容易看出,右端各项的均方导数为:

$$\frac{d}{dt}\left(X_0\exp\left\{\int_{t_0}^{t}a(\tau)d\tau\right\}\right)$$
$$=X_0\exp\left\{\int_{t_0}^{t}a(\tau)d\tau\right\}a(t)$$
$$\frac{d}{dt}\left(\int_{t_0}^{t}\exp\left\{\int_{s}^{t}a(\tau)d\tau\right\}Y(s)ds\right)$$
$$=\exp\left\{\int_{s}^{t}a(\tau)d\tau\right\}Y(t)$$
$$+\int_{t_0}^{t}\exp\left\{\int_{s}^{t}a(\tau)d\tau\right\}a(t)Y(s)ds$$
$$=Y(t)+a(t)\int_{t_0}^{t}\exp\left\{\int_{s}^{t}a(\tau)d\tau\right\}Y(s)ds$$

于是有

$$X'(t)=a(t)\left[X_0\exp\left\{\int_{t_0}^{t}a(\tau)d\tau\right\}\right.$$

$$\left.+\int_{t_0}^{t}\exp\left\{\int_{s}^{t}a(\tau)d\tau\right\}Y(s)ds\right]+Y(t)$$

$$=a(t)X(t)+Y(t)$$

定理证毕.

§3. 正 态 过 程

在本节中，我们将假定所研究的二阶矩过程$\{X(t),t\in T\}$同时还是实的正态过程．如第一章所述，这时由它的均值函数和协方差函数$\Gamma(s,t)=E\{X(s)X(t)\}$就可以完全决定$\{X(t),\ t\in T\}$的有限维分布，从而不仅可以在二阶矩的范围内，而且可以在分布的范围内讨论各种问题，并且得到不少更强的结果.

一、正态过程的导数和积分的分布

在许多实际问题中，往往需要考虑正态过程的导数和积分的分布，为此我们先证明下列定理.

定理 1　若$X^{(n)}=(X_1^{(n)},X_2^{(n)},\cdots,X_k^{(n)})^{\tau}$①为$k$维实正态随机变量，又$X^{(n)}$均方收敛于$X=(X_1,\cdots,X_k)^{\tau}$，即对每个$i$有

$$\lim_{n\to\infty}E|X_i^{(n)}-X_i|^2=0,\qquad 1\leqslant i\leqslant k$$

则X也是k维正态变量.

［证明］　若$X^{(n)}$，X的均值向量和协方差阵分别记为

$$EX^{(n)}=\mu^{(n)}=(\mu_1^{(n)},\mu_2^{(n)},\cdots,\mu_k^{(n)})^{\tau}$$

$$EX=\mu=(\mu_1,\cdots,\mu_k)^{\tau}$$

$$E\{(X^{(n)}-\mu^{(n)})(X^{(n)}-\mu^{(n)})^{\tau}\}=\Sigma^{(n)}=(\sigma_{ij}^{(n)})$$

$$E\{(X-\mu)(X-\mu)^{\tau}\}=\Sigma=(\sigma_{ij})$$

① (　　)$^{\tau}$表示向量或矩阵的共轭转置.

则由定理 1.7 和 1.8 有

$$\lim_{n\to\infty}\mu_j^{(n)}=\mu_j, \qquad 1\leqslant j\leqslant k \tag{1}$$

$$\lim_{n\to\infty}\sigma_{lj}^{(n)}=\sigma_{lj}, \qquad 1\leqslant l, j\leqslant k \tag{2}$$

若以 $\varphi_n(u)$ 和 $\varphi(u)$ 分别表示 $X^{(n)}$ 和 X 的特征函数，则按 $X^{(n)}$ 为 k 维正态分布，必有

$$\varphi_n(u)=\exp\left\{iu'\mu^{(n)}-\frac{1}{2}u'\Sigma^{(n)}u\right\}$$

又由(1)，(2)可得

$$\lim_{n\to\infty}\varphi_n(u)=\exp\{iu'(\lim_{n\to\infty}\mu^{(n)})-\frac{1}{2}u'(\lim_{n\to\infty}\Sigma^{(n)})u\}$$

$$=\exp\{iu'\mu-\frac{1}{2}u'\Sigma u\} \tag{3}$$

另一方面，由定理 12 系 2 可得

$$\lim_{n\to\infty}\varphi_n(u)=\varphi(u)$$

比较上式和(3)可得

$$E\{e^{iu'X}\}=\varphi(u)=\exp\{iu'\mu-\frac{1}{2}u'\Sigma u\}$$

所以 X 也是 k 维正态随机变量.

定理 2 设 $X_T=\{X(t), t\in T\}$ 是正态过程，且在 T 上均方可微，则 $\{X'(t), t\in T\}$ 也是正态过程.

［证明］ 对任意的 $t_1, t_2, \cdots, t_k\in T$ 及 h，利用多维正态变量对线性变换的不变性可得

$$\frac{X(t_1+h)-X(t_1)}{h}, \frac{X(t_2+h)-X(t_2)}{h}, \cdots, \frac{X(t_k+h)-X(t_k)}{h}$$

也是 k 维正态变量. 由于 $X(t)$ 均方可微，所以对每个 j，$\frac{1}{h}(X(t_j+h)-X(t_j))$ 均方收敛于 $X'(t_j)$，利用定理 1 可得 $(X'(t_1),\cdots,X'(t_k))'$ 是 k 维正态变量，所以 $\{X'(t), t\in T\}$ 是正态过程.

系 对正态过程 X_T，若 $EX(t)=m(t)$，$\operatorname{cov}(X(s),X(t))=\Gamma(s,t)$，且 X_T 均方可微，则 $(X'(t_1),\cdots,X'(t_k))'$ 的特征函数为

$$\varphi_{X'}(u_1,t_1;u_2,t_2;\cdots;u_k,t_k)$$

$$=\exp\left\{i\sum_{j=1}^{k}u_jm'(t_j)-\frac{1}{2}\sum_{j,l=1}^{k}u_ju_l\frac{\partial^2\Gamma(t_j,t_l)}{\partial s\partial t}\right\} \tag{4}$$

其中

$$\frac{\partial^2\Gamma(t_j,t_l)}{\partial s\partial t}=\frac{\partial^2\Gamma(s,t)}{\partial s\partial t}\bigg|_{s=t_j,t=t_l}$$

定理 3 设 $X_T=\{X(t),\ t\in T\}$ 是 T 上均方可积的正态过程，则

$$Y(t)=\int_a^t X(s)ds,\qquad a,t\in T$$

也是正态过程.

[证明] 对任一 $t_j\in T$，则按均方积分的定义，$Y(t_j)$ 可表为下列形式的 $Y_j^{(n)}$ 的均方极限，

$$Y_j^{(n)}=\sum_{l=1}^{n_j}X(s_l)\Delta s_l$$

其中 $a=s_0<s_1<\cdots<s_{n_j}$ 是 $[a,t_j]$ 上的一列分点，由于 X_T 为正态过程，所以表为 X_T 线性组合的 $(Y_1^{(n)},\cdots,Y_k^{(n)})'$，是 k 维正态变量，因此按定理 1 它们的均方极限 $(Y(t_1),\cdots,Y(t_k))'$ 也是正态变量，即 Y_T 为正态过程.

系 对正态过程 $X_T=\{X(t),t\in T\}$，若

$$E\{X(t)\}=m(t)$$

$$\operatorname{cov}(X(t),X(s))=\Gamma(t,s)$$

又 X_T 均方可积，

$$Y(t)=\int_a^t X(s)ds$$

则 $(Y(t_1),\cdots,Y(t_k))'$ 的特征函数为

$$
\begin{aligned}
&\varphi_Y(u_1, t_1; u_2, t_2; \cdots; u_k, t_k) \\
&= \exp\left\{ i \sum_{j=1}^{k} u_j \int_a^{t_j} m(s)\, ds \right. \\
&\left. -\frac{1}{2} \sum_{j,l=1}^{k} u_j u_l \int_0^{t_j}\int_0^{t_l} \Gamma(s, t)\, ds dt \right\} \qquad (5)
\end{aligned}
$$

二、零交与阈交问题①

本段中要研究在工程设计中的一个重要问题. 设 $X(t), t\in T$ 是一个随机过程, 我们希望知道在某一时间间隔$[t_1, t_2]\subset T$ 中, $X(t)$的零点个数, 或者更一般地, $X(t)$与某一水平线$x=x_0$ 的交点个数(参看图 1, 其中 $X(t)$是某次观察的结果), 记为$N(x_0; t_1, t_2)$.

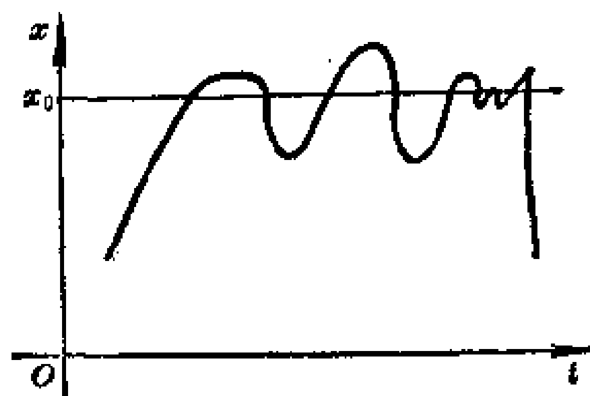


图 1

由于 $X(t)$ 的随机性, 所以 $N(x_0; t_1, t_2)$ 也是随机的, 对固定的 x_0, t_1, t_2, 它是一个随机变量. 我们希望知道它的分布, 或者至少能知道它的数字特征, 如均值等.

如果用 $X(t)$表示时刻 t 某结构中的应力(例如飞机飞行时机翼中的应力或某高层建筑物中的应力等), 则由于各种复杂的因素, 它一般是随机的, 即 $X(t)$是随机过程. 再用 x_0 表示此结构所用材料的疲劳应力, 则不少场合是需要根据应力超过疲劳应力的

---

① 本段只是对零交与阈交问题基本结果的一个直观介绍, 对这一问题的严格处理可参阅 H. Cramer, M. R. Leadbetter, Stationary and Related Stochastic Processes, (1967).

重复次数来进行结构设计. 在数学上也就是要求出 $N(x_0; t_1, t_2)$ 的分布或其均值等数字特征.

这个问题一般称为**零交问题**($x_0=0$ 时)和**阈交问题**(对一般的 x_0), 它最早为美国学者拉爱斯(S. O. Rice)在四十年代所研究.

今后设 $X(t)$, $t\in T$ 是均方可微的正态过程. 引入下列新过程可使阈交问题更便于处理:

$$Y(t)=u[X(t)-x_0],\quad t\in T$$

其中 $u(x)$ 为单位阶梯函数, 即

$$u(x)=\begin{cases}1, & x\geqslant 0\\ 0, & x<0\end{cases}$$

对 $Y(t)$ 求导数, 形式上有

$$Y'(t)=X'(t)\cdot\delta[X(t)-x_0]$$

其中 $\delta(x)$ 是狄拉克的 δ 函数, 它是一个广义函数. 直观上可理解为在原点的一个无限窄又无限高的尖脉冲, 但要求其下的面积为 1, 即

$$\int_{-\infty}^{\infty}\delta(x)dx=1$$

例如可以看成是 $\frac{1}{\sqrt{2\pi}\,\sigma}e^{-\frac{x^2}{2\sigma^2}}$, 当 $\sigma\to+0$ 时的极限.

$\delta(x)$ 有如下的一个重要性质. 对任意函数 $g(x)$, 有

$$g(x)=\int_{-\infty}^{\infty}\delta(y)g(x-y)dy \tag{6}$$

直观上, 上式是容易理解的, 因为 $\delta(x)$ 在 $x\neq 0$ 处都是零, 因而右边的积分

$$\int_{-\infty}^{\infty}\delta(y)g(x-y)dy=\int_{-\varepsilon_1}^{\varepsilon_2}\delta(y)g(x-y)dy$$

$$\doteq g(x-\tilde{y})\int_{-\varepsilon_1}^{\varepsilon_2}\delta(y)dy=g(x-\tilde{y}),\quad -\varepsilon_1\leqslant\tilde{y}\leqslant\varepsilon_2$$

由于 ε_1, ε_2 可以任意小, 所以就得到(6)式.

为了帮助理解上述过程，我们将 $X(t)$，$Y(t)$和$Y'(t)$的某个现实或样本函数画成图 2.

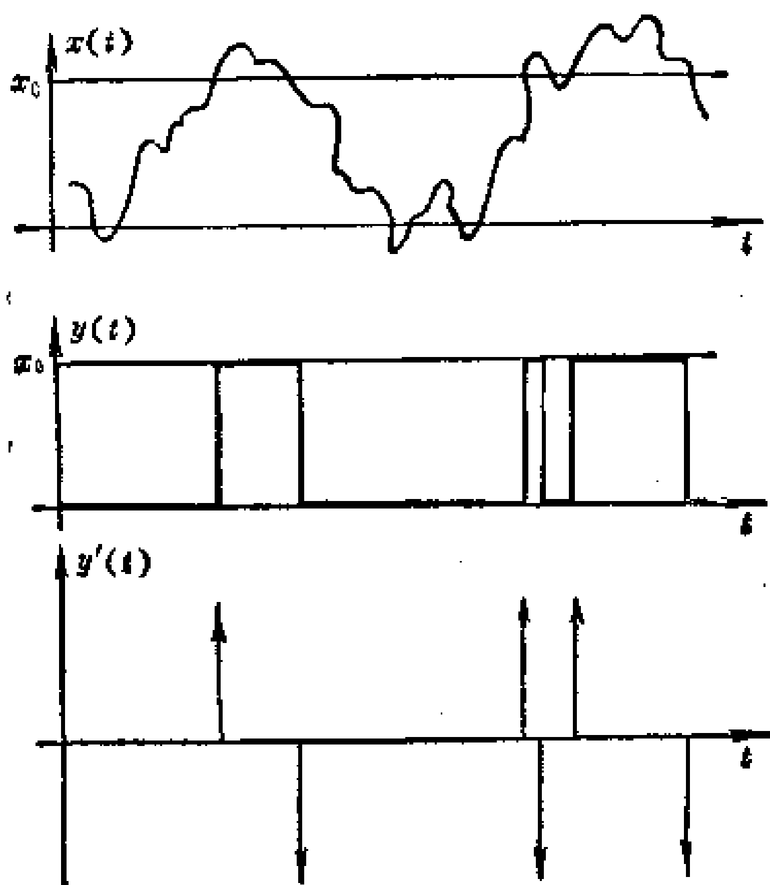


图 2

由图 2 可以看出，$Y'(t)$的样本函数由一些 δ 函数，或一些单位脉冲所组成，其方向向上或向下，则视 $X(t)$以正的斜率还是以负的斜率与水平线 x_0 相交而定. 同时，由图 2 容易看出：

$$N(x_0; t_1, t_2) = \int_{t_1}^{t_2} |Y'(t)| dt$$

$$= \int_{t_1}^{t_2} |X'(t)| \delta[X(t) - x_0] dt$$

两边取数学期望，则得

$$E\{N(x_0; t_1, t_2)\}$$

$$= \int_{t_1}^{t_2} E\{|X'(t)| \delta[X(t) - x_0]\} dt$$

如果用$f(t,x;t,x')$表示$X(t)$和$X'(t)$的联合分布密度，则有

$$E\{N(x_0;t_1,t_2)\}$$
$$=\int_{t_1}^{t_2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}|x'|\delta(x-x_0)f(t,x;t,x')dxdx'dt$$

根据(6)式，有

$$E\{N(x_0;t_1,t_2)\}$$
$$=\int_{t_1}^{t_2}\int_{-\infty}^{\infty}|x'|f(t,x_0;t,x')dx'dt \tag{7}$$

如记

$$r(t,x_0)=\int_{-\infty}^{\infty}|x'|f(t,x_0;t,x')dx' \tag{8}$$

则由(7)式可知，$r(t,x_0)$是$X(t)$在t时与x_0相交的平均率. 特别在$X(t)$, $t\in T$是平稳时，$r(t,x_0)$与t无关，这时$r(t,x_0)=r(x_0)$，它表示单位时间中，$X(t)$与x_0相交的次数.

类似地，如用$f(\tau_1,x_1;\tau_1,x_1';\tau_2,x_2;\tau_2,x_2')$表示$X(\tau_1)$,$X'(\tau_1)$,$X(\tau_2)$, $X'(\tau_2)$的联合分布密度，则可得

$$E\{(N(x_0;t_1,t_2))^2\}$$
$$=\int_{t_1}^{t_2}\int_{t_1}^{t_2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}|x_1'|\cdot|x_2'|\cdot f(\tau_1,x_0;\tau_1,x_1';\tau_2,x_0;\tau_2,x_2')dx_1'dx_2'd\tau_1d\tau_2$$

等等.

在有些场合，我们只关心向上与x_0相交的个数和向下与x_0相交的个数，如分别记它们为$N_+(x_0;t_1,t_2)$和$N_-(x_0;t_1,t_2)$，则类似地有：

$$E\{N_+(x_0;t_1,t_2)\}=\int_{t_1}^{t_2}\int_0^{\infty}x'f(t,x_0;t,x')dx'dt$$

$$E\{N_-(x_0;t_1,t_2)\}=-\int_{t_1}^{t_2}\int_{-\infty}^{0}x'f(t,x_0;t,x')dx'dt$$

等等.

为了求得简单的表达式，我们进一步假定 $X(t)$是平稳的. 此时我们有

引理 设 $X(t)$是平稳的均方可微的实二阶矩过程，则

$$E\{X(t)X'(t)\}=0$$

［证明］ 由于 $X(t)$是平稳的，所以

$$\Gamma(s,t)=B(s-t)$$

因为 $\Gamma(s,t)$是对称的，

$$\Gamma(s,t)=\Gamma(t,s)$$

所以

$$B(\tau)=B(-\tau)$$

由于 $X(t)$均方可微，$\Gamma(s,t)$广义二次可导，因而 $B(\tau)$二次可导. 对上式求导，得

$$B'(\tau)=-B'(-\tau)$$

特别取 $\tau=0$，得

$$B'(0)=-B'(0)$$

即

$$B'(0)=0$$

但由定理 2.2 的系 2 知

$$E\{X(t)X'(t)\}=\frac{\partial}{\partial t}\Gamma(t,t)=B'(0)=0$$

系 如 $X(t)$还是正态的，则 $X(t)$和 $X'(t)$独立.

由上述定理及其系可知，对平稳正态均方可微的二阶矩过程 $X(t)$，$t\in T$，$X'(t)$和 $X(t)$是相互独立的，于是 $X(t)$和 $X'(t)$的联合分布密度是

$$\begin{aligned}&f(t,x;t,x')=f(x,x')\\&=\frac{1}{\sqrt{2\pi}\sigma_X}\exp\left\{-\frac{1}{2}\frac{x^2}{\sigma_X^2}\right\}\frac{1}{\sqrt{2\pi}\sigma_{X'}}\exp\left\{-\frac{1}{2}\frac{x'^2}{\sigma_{X'}^2}\right\}\end{aligned}$$

$$=\frac{1}{2\pi\sigma_X\sigma_{X'}}\exp\left\{-\frac{1}{2}\left(\frac{x^2}{\sigma_X^2}+\frac{x'^2}{\sigma_{X'}^2}\right)\right\}$$

其中

$$\sigma_X^2=\Gamma(t,\ t)=B(0)$$

$$\sigma_{X'}^2=\frac{\partial^2}{\partial s\partial t}\Gamma(s,\ t)\Big|_{s=t=t}=-B''(0)$$

按(8)式，有

$$r(x_0)=\int_{-\infty}^{\infty}|x'|f(x_0,x')dx'$$

$$=\int_{-\infty}^{\infty}|x'|\frac{1}{2\pi\sigma_X\sigma_{X'}}\exp\left\{-\frac{1}{2}\left(\frac{x_0^2}{\sigma_X^2}+\frac{x'^2}{\sigma_{X'}^2}\right)\right\}dx'$$

$$=\frac{1}{2\pi\sigma_X\sigma_{X'}}\exp\left\{-\frac{1}{2}\cdot\frac{x_0^2}{\sigma_X^2}\right\}$$

$$\cdot\int_{-\infty}^{\infty}|x'|\exp\left\{-\frac{1}{2}\cdot\frac{x'^2}{\sigma_{X'}^2}\right\}dx'$$

$$=\frac{\sigma_{X'}}{\pi\sigma_X}\exp\left\{-\frac{1}{2}\cdot\frac{x_0^2}{\sigma_X^2}\right\}\tag{9}$$

因此得

定理 4 设 $X(t)$, $t\in T$ 是实的均方可微的平稳正态过程，有

$$r(x_0)=\frac{\sigma_{X'}}{\pi\sigma_X}\exp\left\{-\frac{1}{2}\frac{x_0^2}{\sigma_X^2}\right\}$$

中

$$\sigma_X^2=B(0),\ \sigma_{X'}^2=-B''(0)$$

$B(\tau)$是 $X(t)$的协方差函数。

三、正态马尔可夫过程

本节中我们给出正态过程 $X(t)$, $t\in T$ 同时又是马尔可夫过程的一个充要条件. 为此先引入规范化的**相关函数**

$$R(s,t)=\begin{cases}\dfrac{\Gamma(s,\ t)}{\sqrt{\Gamma(s,\ s)\Gamma(t,\ t)}}, & \text{如 } \Gamma(s,\ s)\Gamma(t,\ t)>0\\ 0, & \text{如 } \Gamma(t,\ t)=0\end{cases}$$

其中 $\Gamma(s, t)$是 $X(t), t\in T$ 的相关函数.

我们有:

定理 5 设 $X(t), t\in T$ 是正态过程, 且 $E|X(t)|^2\neq 0$, 则它同时又是马尔可夫过程的充要条件是其规范化的相关函数满足:

$$R(t_1, t_3)=R(t_1, t_2)R(t_2, t_3), t_1<t_2<t_3.$$

[证明] 由于要用到一般的条件数学期望等概念, 这儿从略①.

§4. 伊藤随机积分和伊藤随机微分方程

本节中将研究如下的随机积分

$$\int_a^b X(t)dW(t)$$

其中不仅 $X(t)$是实的二阶矩过程, 而且$W(t)$也是实的二阶矩过程. 但由于这里的情形比 §2 中要复杂得多, 所以, 我们限于$W(t)$是第一章中的维纳过程的场合. 这种随机积分最初由日本学者伊藤开始研究, 所以就称为伊藤随机积分, 它在近代滤波, 通讯, 控制理论中都有广泛的应用.

一、维纳过程及其形式导数

今后用$W(t), t\geqslant 0$表示第一章中的维纳过程, 也叫布朗运动. 我们知道, 它的均值函数为零, 协方差函数是

$$\Gamma(s, t)=\sigma^2\cdot\min(s, t)$$

特别有

$$E\{[W(t)-W(s)]^2\}=\sigma^2(t-s),\quad t>s$$

由于 $\Gamma(s, t)$是连续的, 所以$W(t)$是均方连续的, 但由于下列

① 有兴趣的读者可参阅 M. Loeve, "Probability Theory", 第三版(1963 年), 第 464 页.

偏导数不存在，

$$\frac{\partial^2}{\partial s\partial t}\Gamma(s,t)=\sigma^2\frac{\partial}{\partial s}U(s-t)$$

其中

$$U(s-t)=\begin{cases}1, & s>t\\ 0, & s<t\end{cases}$$

因此$W(t)$的均方导数不存在，但形式上，有

$$\Gamma_{W'}(s,t)=\frac{\partial^2}{\partial s\partial t}\Gamma(s,t)=\sigma^2\cdot\delta(s-t)$$

其中$\delta(\cdot)$ 是狄拉克的δ函数．它形式上表示$W'(s)$和$W'(t)$ 当$s\neq t$ 时相关为零，即不相关．这种随机过程一般称为**白噪声**．

由于$W(t)$是正态过程，所以它的形式导数$W'(t)$形式地称为正态的白噪声．

不失一般性，今后假定$\sigma^2=1$．

二、伊藤随机积分

现在来考虑

$$\int_a^b X(t)dW(t),\ b>a\geqslant 0$$

的定义．引入$[a,b]$的一组分点

$$a=t_0<t_1<\cdots<t_n=b$$

$$\Delta_n=\max_{1\leqslant k\leqslant n}(t_k-t_{k-1})$$

并考察和式

$$Y_n=\sum_{k=1}^{n}X(t_k')[W(t_k)-W(t_{k-1})]$$

其中$t_{k-1}\leqslant t_k'\leqslant t_k$, $1\leqslant k\leqslant n$．如果在$\Delta_n\to 0$时，Y_n均方收敛，则自然将此极限定义为上述积分．但可惜的是，当t_k'随意取时，极限并不存在．因此，在这种意义下，上述的随机积分并不存在．伊藤代替Y_n，而考察下列和式，

$$I_n=\sum_{k=1}^{n}X(t_{k-1})[W(t_k)W-(t_{k-1})]$$

即 t_k' 确定地取为左端点 t_{k-1}，而不是可以在 $[t_{k-1}, t_k]$ 中任意选取. 如果在 $\Delta_n\to 0$ 时, I_n 均方收敛, 则其极限称为 $X(t)$ 关于维纳过程 $W(t)$ 的伊藤积分, 记为

$$(I)\int_a^b X(t)dW(t)=\lim_{\Delta_n\to 0}I_n$$

顺便指出,苏联学者斯特拉塔诺维奇 (Стратонович) 考虑下列和式:

$$C_n=\sum_{k=1}^{n}X\left(\frac{t_{k-1}+t_k}{2}\right)[W(t_k)-W(t_{k-1})]$$

即 t_k' 确定地取为 $[t_{k-1}, t_k]$ 的中点, 并在 $\Delta_n\to 0$ 时, 当 C_n 均方收敛时, 称其极限为 $X(t)$ 关于维纳过程 $W(t)$ 的随机积分, 记为

$$(C)\int_a^b X(t)dW(t)$$

这二种积分的定义是十分相近的, 由于在一般的文献中伊藤积分应用得更为广泛, 所以我们今后只限于讨论伊藤积分, 而且一般均略去前面的 (I).

我们有如下的存在定理.

定理 1 设 $X(t)$ 是均方连续的二阶矩过程, 并且对任意的 $s_1', s_2'\leqslant t_{k-1}<t$ 及 $s_1<s_2\leqslant t_{k-1}$, $(X(s_1'), X(s_2'), W(s_2)-W(s_1))$ 与 $W(t_k)-W(t_{k-1})$ 相互独立, 则 $X(t)$ 关于 $W(t)$ 的伊藤积分存在且唯一.

[证明] 根据定理 1.11, 我们只要证明下式极限存在即可。

$$E\{I_mI_n\}=\sum_{j=1}^{m}\sum_{k=1}^{n}E\{X(s_{j-1})X(t_{k-1})[W(s_j)-W(s_{j-1})]\cdot[W(t_k)-W(t_{k-1})]\}$$

其中

$$a=s_0<s_1<\cdots<s_m=b$$

$$a=t_0<t_1<\cdots<t_n=b$$

是$[a, b]$的两个分法.

上式右边各项可根据长方形$[s_{j-1}, s_j]\times[t_{k-1}, t_k]$是否与正方形$[a, b]\times[a, b]$的主对角线相交而分为两类：第一类是相交的，第二类是不相交的(参看图 1).

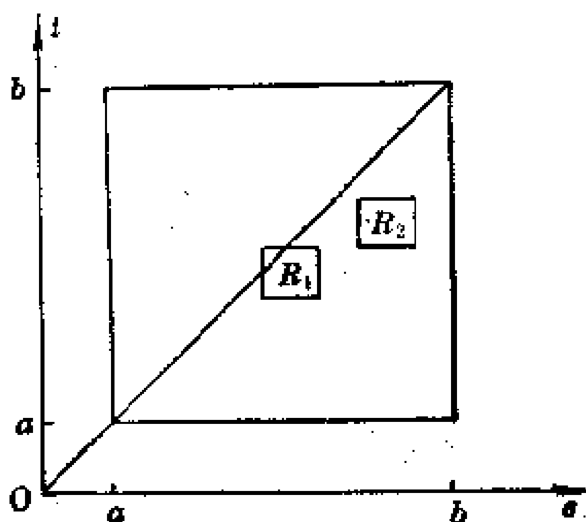


图 1

先看第二类的项，此时有$s_{j-1}<s_j\leqslant t_{k-1}<t_k$(或者$t_{k-1}<t_k\leqslant s_{j-1}<s_j$)，于是由定理假定知，$X(s_{j-1})$，$X(t_{k-1})$，$W(s_j)-W(s_{j-1})$与$W(t_k)-W(t_{k-1})$相互独立的(或者$X(s_{j-1})$，$X(t_{k-1})$，$W(t_k)-W(t_{k-1})$与$W(s_j)-W(s_{j-1})$是相互独立的)，于是有

$$\begin{aligned}
&E\{X(s_{j-1})X(t_{k-1})[W(s_j)-W(s_{j-1})][W(t_k)-W(t_{k-1})]\}\\
&\quad=E\{X(s_{j-1})X(t_{k-1})[W(s_j)-W(s_{j-1})]\}\\
&\qquad\cdot E\{W(t_k)-W(t_{k-1})\}\\
&\quad=E\{X(s_{j-1})X(t_{k-1})[W(s_j)-W(s_{j-1})]\}\cdot 0\\
&\quad=0
\end{aligned}$$

显然可见，当$t_{k-1}<t_k\leqslant s_{j-1}<s_j$时，上式也成立，因此，所有第二类项都是零.

再看第一类项，此时$[s_{j-1}, s_j]$和$[t_{k-1}, t_k]$有公共部分，一种典型的场合是：$s_{j-1}<t_{k-1}<s_j<t_k$(其它场合可以类似地加以分析). 我们有

$$\begin{aligned}
&E\{X(s_{j-1})X(t_{k-1})[W(s_j)-W(s_{j-1})][W(t_k)-W(t_{k-1})]\}\\
&=E\{X(s_{j-1})X(t_{k-1})[(W(s_j)-W(t_{k-1}))+(W(t_{k-1})\\
&\quad-W(s_{j-1}))][(W(t_k)-W(s_j))+(W(s_j)-W(t_{k-1})]\}
\end{aligned}$$

$$
\begin{aligned}
&= E\{X(s_{j-1})X(t_{k-1})[W(s_j)-W(t_{k-1})][W(t_k)-W(s_j)]\} \\
&\quad + E\{X(s_{j-1})X(t_{k-1})[W(s_j)-W(t_{k-1})][W(s_j)-W(t_{k-1})]\} \\
&\quad + E\{X(s_{j-1})X(t_{k-1})[W(t_{k-1})-W(s_{j-1})] \\
&\qquad [W(t_k)-W(s_j)]\} \\
&\quad + E\{X(s_{j-1})X(t_{k-1})[W(t_{k-1})-W(s_{j-1})] \\
&\qquad [W(s_j)-W(t_{k-1})]\}
\end{aligned}
$$

仿上面关于第二类项的讨论，可以知道，上述四项中的第一、三、四项都是零，剩下的第二项等于

$$
\begin{aligned}
&E\{X(s_{j-1})X(t_{k-1})\}\cdot E\{[W(s_j)-W(t_{k-1})]^2\} \\
&= \Gamma(s_{j-1}, t_{k-1})\cdot(s_j-t_{k-1})
\end{aligned}
$$

其中 $\Gamma(s, t)$是 $X(t)$的协方差函数，且如前所述，一般取$W(t)$的 $\sigma^2=1$(显然这样做并不失去一般性)．于是对第一类项有

$$
\begin{aligned}
&E\{X(s_{j-1})X(t_{k-1})[W(s_j)-W(s_{j-1})] \\
&\quad [W(t_k)-W(t_{k-1})]\} \\
&= \Gamma(s_{j-1}, t_{k-1})\cdot(s_j-t_{k-1})
\end{aligned}
$$

其中 s_j-t_{k-1} 是$[s_{j-1}, s_j]$和$[t_{k-1}, t_k]$公共部分的长度．

若将两组分点$a=s_0<s_1<\cdots<s_m=b$及$a=t_0<t_1<\cdots<t_n=b$合并为一组，并以 $a_0=u_0<u_1<\cdots<u_{m+n-1}=b$ 表这些分点（若有 s_i, t_j 相重时，u_l的个数将小于 $m+n$)，则可得

$$
E\{I_m I_n\} = \sum_{l=1}^{m+n-1} \{\Gamma(u_{l-1}, u_{l-1})(u_l-u_{l-1}) + o(u_l-u_{l-1})\}
$$

由于 $\Gamma(t, t)$是连续的，因而是可积的，所以有

$$
\lim E\{I_m I_n\} = \int_a^b \Gamma(t, t)\,dt
$$

存在性于是得证．唯一性由定理 1.10 直接推出．

系 在定理的条件下，如相应的$W(t)$的 $\sigma^2=1$，则有

$$
E\left\{\left(\int_a^b X(t)\,dW(t)\right)^2\right\} = \int_a^b E\{X^2(t)\}\,dt
$$

即 $$\left\|\int_a^b X(t)dW(t)\right\|^2=\int_a^b \|X(t)\|^2dt$$

下面来看一个简单的例子．试求

$$\int_a^b W(t)dW(t)$$

其中$W(t)$的 $\sigma^2=1$．这儿的被积过程 $W(t)$显然满足定理的条件，因此上述积分是存在的，取$[a, b]$的一组分点

$$a=t_0<t_1<\cdots<t_n=b,\quad \Delta_n=\max_{1\leqslant k\leqslant n}(t_k-t_{k-1})$$

作下列和式

$$\begin{aligned}
&\sum_{k=1}^n W(t_{k-1})[W(t_k)-W(t_{k-1})]\\
&=-\sum_{k=1}^n W(t_{k-1})[W(t_{k-1})-W(t_k)]\\
&=-[W(t_0)^2-W(t_0)W(t_1)+W(t_1)^2-W(t_1)W(t_2)\\
&\quad+\cdots+W(t_{n-1})^2-W(t_{n-1})W(t_n)]\\
&=-\frac{1}{2}\{W(t_0)^2+\sum_{k=1}^n[W(t_k)-W(t_{k-1})]^2-W(t_n)^2\}\\
&=\frac{1}{2}[W(b)^2-W(a)^2]-\frac{1}{2}\sum_{k=1}^n[W(t_k)-W(t_{k-1})]^2
\end{aligned}$$

于是有

$$\begin{aligned}
&\int_a^b W(t)dW(t)\\
&=\frac{1}{2}[W(b)^2-W(a)^2]-\frac{1}{2}\lim_{\Delta_n\to 0}\sum_{k=1}^n[W(t_k)-W(t_{k-1})]^2
\end{aligned}$$

现在来计算右边的极限，为简化记号，以下令

$$W(t_k)-W(t_{k-1})=\Delta W_k$$

$$t_k-t_{k-1}=\Delta t_k$$

则我们有

$$\left\|\sum_{k=1}^{n}\Delta W_k^2-(b-a)\right\|^2=\left\|\sum_{k=1}^{n}(\Delta W_k^2-\Delta t_k)\right\|^2$$

$$=\sum_{k=1}^{n}\|\Delta W_k^2-\Delta t_k\|^2$$

$$+2\sum_{\substack{i,j\\i\neq j}}E\{(\Delta W_i^2-\Delta t_i)(\Delta W_j^2-\Delta t_j)\}$$

$$=\sum_{k=1}^{n}\|\Delta W_k^2-\Delta t_k\|^2+2\sum_{\substack{i,j\\i\neq j}}E\{\Delta W_i^2-\Delta t_i\}E\{\Delta W_j^2-\Delta t_j\}$$

$$=\sum_{k=1}^{n}\|\Delta W_k^2-\Delta t_k\|^2$$

$$=\sum_{k=1}^{n}E\{\Delta W_k^4-2\Delta W_k^2\cdot\Delta t_k+\Delta t_k^2\}$$

$$=\sum_{k=1}^{n}\{3\Delta t_k^2-2\Delta t_k^2+\Delta t_k^2\}$$

$$=2\sum_{k=1}^{n}\Delta t_k^2\leqslant 2\Delta_n\cdot\sum_{k=1}^{n}\Delta t_k=2\Delta_n(b-a)\to 0$$

即

$$\lim_{\Delta_n\to 0}\sum_{k=1}^{n}[W(t_k)-W(t_{k-1})]^2=b-a$$

于是综上所述，可得：

$$\int_a^b W(t)dW(t)=\frac{1}{2}[W(b)^2-W(a)^2]-\frac{1}{2}(b-a)$$

从这个例子，我们看到伊藤随机积分与普通的积分不一样．对它通常的分部积分公式不成立．事实上，如果$W(t)$是普通的函数，则右边只有第一项，而没有第二项$\frac{1}{2}(b-a)$．这一点在进行伊藤积分的运算时，是需要特别注意的．

在上述例子讨论的过程中，我们证明了

$$\lim_{\Delta_n\to 0}\sum_{k=1}^{n}[W(t_k)-W(t_{k-1})]^2=b-a \tag{1}$$

这是关于维纳过程 $W(t)$的一个十分重要的事实．下面我们利用这一事实来说明伊藤积分与我们前面讨论的均方积分不同，即积分和式中的 t_k' 不能任意选取．事实上，设

$$a=t_0<t_1<t_2<\cdots<t_n=b$$

$$\Delta_n=\max_{1\leqslant k\leqslant n}(t_k-t_{k-1})$$

是$[a,b]$的一组分点，考虑下列两个和式

$$I_n=\sum_{k=1}^{n}W(t_{k-1})[W(t_k)-W(t_{k-1})]$$

和

$$J_n=\sum_{k=1}^{n}W(t_k)[W(t_k)-W(t_{k-1})]$$

则它们形式上都是

$$\int_a^b W(t)dW(t)$$

的积分和式，分别相应于 $t_k'=t_{k-1}$ 和 $t_k'=t_k$．两式相减得

$$J_n-I_n=\sum_{k=1}^{n}[W(t_k)-W(t_{k-1})]^2\to b-a\quad(\Delta_n\to 0)$$

因此 I_n 和 J_n 有不同的均方极限，即积分和式中的 t_k' 是不能任意取的，于是通常意义下的均方积分不存在．

上面已经说过伊藤积分没有通常的分部积分公式，而 t_k' 的选择又直接影响到积分的定义，那么能否选择 t_k'的特殊取法而让规定的积分具有通常的分部积分公式呢，下面我们来分析这一点．如前所述，I_n的均方极限是存在的，

$$\begin{aligned}\lim_{\Delta_n\to 0}I_n&=\int_a^b W(t)dW(t)\\&=\frac{1}{2}[W(b)^2-W(a)^2]-\frac{1}{2}(b-a)\end{aligned}$$

所以 $J_n=I_n+(J_n-I_n)$的均方极限也存在，且有

$$\lim_{\Delta_n\to 0} J_n=\frac{1}{2}[W(b)^2-W(a)^2]-\frac{1}{2}(b-a)+(b-a)$$

$$=\frac{1}{2}[W(b)^2-W(a)^2]+\frac{1}{2}(b-a)$$

于是

$$\lim_{\Delta_n\to 0}\frac{I_n+J_n}{2}=\lim_{\Delta_n\to 0}\sum_{k=1}^{n}\frac{W(t_{k-1})+W(t_k)}{2}[W(t_k)-W(t_{k-1})]$$

$$=\frac{1}{2}[W(b)^2-W(a)^2]$$

我们还有

$$\frac{I_n+J_n}{2}-\sum_{k=1}^{n}W\left(\frac{t_{k-1}+t_k}{2}\right)[W(t_k)-W(t_{k-1})]$$

$$=\sum_{k=1}^{n}\left[\frac{W(t_{k-1})+W(t_k)}{2}-W\left(\frac{t_{k-1}+t_k}{2}\right)\right][W(t_k)-W(t_{k-1})]$$

$$=\frac{1}{2}\sum_{k=1}^{n}\left\{\left[W(t_k)-W\left(\frac{t_{k-1}+t_k}{2}\right)\right]-\left[W\left(\frac{t_{k-1}+t_k}{2}\right)-W(t_{k-1})\right]\right\}\cdot\left\{\left[W(t_k)-W\left(\frac{t_{k-1}+t_k}{2}\right)\right]+\left[W\left(\frac{t_{k-1}+t_k}{2}\right)-W(t_{k-1})\right]\right\}$$

$$=\frac{1}{2}\sum_{k=1}^{n}\left\{\left[W(t_k)-W\left(\frac{t_{k-1}+t_k}{2}\right)\right]^2-\left[W\left(\frac{t_{k-1}+t_k}{2}\right)-W(t_{k-1})\right]^2\right\}$$

注意到诸区间$\left[t_{k-1},\frac{t_{k-1}+t_k}{2}\right]$，$1\leqslant k\leqslant n$ 长度之和以及诸区间$\left[\frac{t_{k-1}+t_k}{2},t_k\right]$，$1\leqslant k\leqslant n$长度之和都是$\frac{1}{2}(b-a)$，仿(1)式的证明，

可知

$$\lim_{\Delta_n\to 0}\sum_{k=1}^{n}\left[W(t_k)-W\left(\frac{t_{k-1}+t_k}{2}\right)\right]^2=\frac{1}{2}(b-a)$$

和

$$\lim_{\Delta_n\to 0}\sum_{k=1}^{n}\left[W\left(\frac{t_{k-1}+t_k}{2}\right)-W(t_{k-1})\right]^2=\frac{1}{2}(b-a)$$

因此有

$$\lim_{\Delta_n\to 0}\frac{I_n+J_n}{2}-\sum_{k=1}^{n}W\left(\frac{t_{k-1}+t_k}{2}\right)[W(t_k)-W(t_{k-1})]=0$$

或者

$$\lim_{\Delta_n\to 0}\sum_{k=1}^{n}W\left(\frac{t_{k-1}+t_k}{2}\right)[W(t_k)-W(t_{k-1})]$$

$$=\lim_{\Delta_n\to 0}\frac{I_n+J_n}{2}=\frac{1}{2}[W(b)^2-W(a)^2]$$

若取$t'_k=\dfrac{t_{k-1}+t_k}{2}$，即按斯特拉塔诺维奇方法规定随机积分，则有

$$(C)\int_a^b W(t)dW(t)=\frac{1}{2}[W(b)^2-W(a)^2]$$

它恰好与普通分析中的分部积分形式上一致。

下面再给出伊藤积分的一些性质。

定理 2 设下列伊藤积分存在

$$\int_a^b X(t)dW(t),\ \int_a^b Y(t)dW(t)$$

则对任意常数α, β有

(i) $\int_a^b(\alpha X(t)+\beta Y(t))dW(t)$

$$=\alpha\int_a^b X(t)dW(t)+\beta\int_a^b Y(t)dW(t)$$

(ii) 如$a\leqslant c\leqslant b$，则

$$\int_a^b X(t)dW(t)=\int_a^c X(t)dW(t)+\int_c^b X(t)dW(t)$$

[证明] 很简单，从略.

系 设下列伊藤积分存在:

$$\int_a^b X(t)dW(t)$$

则对 $a\leqslant t\leqslant b$

$$Y(t)=\int_a^t X(s)dW(s)$$

存在且关于 t 是均方连续的.

[证明] $\|Y(t+h)-Y(t)\|^2=\left\|\int_t^{t+h} X(s)dW(s)\right\|^2$

$$=\int_t^{t+h}\|X(s)\|^2ds\to 0 \quad (h\to 0)$$

定理 3 设 $\{X_n(t), t\in T=[a, b]\}$ 是均方连续的二阶矩过程，且满足定理 1 的条件，如果关于 $t\in T$ 一致地有

$$\lim_{n\to\infty} X_n(t)=X(t)$$

则 $X(t)$ 也是均方连续且满足定理 1 的条件，且对一切 $a\leqslant t\leqslant b$，一致地有

$$\lim_{n\to\infty}\int_a^t X_n(s)dW(s)=\int_a^t X(s)dW(s)$$

[证明] 因为 $X_n(t)$ 一致均方收敛于 $X(t)$，$X_n(t)$ 是均方连续的，所以 $X(t)$ 也是均方连续的，且不难证明 $X(t)$ 满足定理 1 的条件.

再由定理 1 可知积分

$$Y_n(t)=\int_a^t X_n(s)dW(s)$$

和

$$Y(t)=\int_a^t X(s)dW(s)$$

都存在，剩下要证明，对 $a\leqslant t\leqslant b$，一致地有

$$\lim_{n\to\infty} Y_n(t)=Y(t)$$

我们有

$$\begin{aligned}\|Y_n(t)-Y(t)\|^2&=\left\|\int_a^t (X_n(s)-X(s))dW(s)\right\|^2\\&=\int_a^t \|X_n(s)-X(s)\|^2 ds\\&\leqslant\int_a^b \|X_n(s)-X(s)\|^2 ds\to 0\end{aligned}$$

其中第二个等式是由于定理 1 的系，而最后一项则与 t 无关，因此收敛关于 t 是一致的，定理证毕。

四、伊藤随机微分方程

在上述伊藤积分的定义基础上，本段中将进一步讨论工程中十分有用的伊藤随机微分方程：

$$\begin{cases}dX(t)=f(t,X(t))dt+g(t,X(t))dW(t),\ t\in T\\X(t_0)=X_0\end{cases}\tag{2}$$

由第一段的讨论可知，如形式上记

$$N(t)=\frac{dW(t)}{dt}$$

则 $N(t)$ 即为白噪声，它是工程中最广泛遇到的噪声的一种简单的数学模型．这时上述方程式也可形式上写为：

$$\begin{cases}\dfrac{dX(t)}{dt}=f(t,X(t))+g(t,X(t))N(t)\\X(t_0)=X_0\end{cases}$$

如没有 $g(t,\ X(t))N(t)$ 这一项，则可视为普通的常微分方程，增加了这一项，则表示引入了随机因素，于是 $X(t)$ 不能再是普通的函数，而必须是随机过程了．

由于普通的常微分方程在许多工程问题中起着重要的作用，因而在更精密分析中不得不考虑随机因素时，随机微分方程就要

起同样重要的作用.

为了避开 $N(t)$ 的奇异性质，代替随机微分方程，我们考虑下列的伊藤积分方程:

$$X(t)=X_0+\int_{t_0}^{t} f(s, X(s))ds+\int_{t_0}^{t} g(s, X(s))dW(s) \quad (3)$$

其中右边第一个积分是均方积分，第二个积分是伊藤随机积分，今后将把(2)和(3)视为等价的.

我们有如下的存在定理.

定理 4 设 $f(t, x)$ 和 $g(t, x)$, $t\in T$, $x\in(-\infty, \infty)$ 是满足下列条件的实的普通函数:

(i) 它们是二元连续函数，而且关于 x 是 t 的一致连续函数;

(ii) 增长条件:

$$|f(t, x)|^2\leqslant K^2(1+x^2)$$

$$|g(t, x)|^2\leqslant K^2(1+x^2)$$

(iii) 李不西兹条件:

$$|f(t, x_1)-f(t, x_2)|\leqslant K|x_2-x_1|$$

$$|g(t, x_1)-g(t, x_2)|\leqslant K|x_2-x_1|$$

其中 K 是某个正数，如 X_0 再与 $\{W(t), t\in T\}$ 独立，则方程(2)有唯一解.

[证明] 先根据定理的条件，证明几个以后要用的不等式. 设 $Y(t)$, $t\in T$ 是二阶矩过程，则有

$$\begin{aligned}\|f(t, Y(t))\|^2&=E\{|f(t, Y(t))|^2\}\\&\leqslant E\{K^2(1+[Y(t)]^2)\}\\&=K^2(1+E\{|Y(t)|^2\})\\&=K^2(1+\|Y(t)\|^2)\end{aligned}$$

类似地有

$$\|g(t, Y(t))\|^2\leqslant K^2(1+\|Y(t)\|^2)$$

由上述两个不等式可知，当 $Y(t)$ 是二阶矩过程时，$f(t, Y(t))$ 和 $g(t, Y(t))$ 也是二阶矩过程。

再设 $Y(t), t\in T$ 是二阶矩过程，则还有

$$\begin{aligned}&\|f(t, Y(t+\Delta t))-f(t, Y(t))\|^2\\&=E\{|f(t, Y(t+\Delta t))-f(t, Y(t))|^2\}\\&\leqslant E\{K^2|Y(t+\Delta t)-Y(t)|^2\}\\&=K^2\|Y(t+\Delta t)-Y(t)\|^2\end{aligned}$$

亦即

$$\begin{aligned}&\|f(t, Y(t+\Delta t))-f(t, Y(t))\|\\&\leqslant K\|Y(t+\Delta t)-Y(t)\|\end{aligned}$$

类似地可证

$$\begin{aligned}&\|g(t, Y(t+\Delta t))-g(t, Y(t))\|\\&\leqslant K\|Y(t+\Delta t)-Y(t)\|\end{aligned}$$

现在我们来利用上述不等式证明，当 $Y(t)$ 均方连续时，$f(t, Y(t))$ 和 $g(t, Y(t))$ 也是均方连续，显见只要对 $f(t, Y(t))$ 证明就够了，我们有

$$\begin{aligned}&\|f(t+\Delta t, Y(t+\Delta t))-f(t, Y(t))\|\\&\leqslant\|f(t+\Delta t, Y(t+\Delta t))-f(t, Y(t+\Delta t))\|\\&+\|f(t, Y(t+\Delta t))-f(t, Y(t))\|\end{aligned}$$

其中第二项，由已证明的不等式有

$$\begin{aligned}&\|f(t, Y(t-\Delta t))-f(t, Y(t)\|\\&\leqslant K\|Y(t+\Delta t)-Y(t)\|\to 0\end{aligned}$$

而第一项为

$$E\{|f(t+\Delta t, Y(t+\Delta t))-f(t, Y(t+\Delta t))|^2\}$$

由于 $f(t, x)$ 关于 x 是 t 的一致连续函数，所以只要 Δt 充分小，即对一切 x 有

$$|f(t+\Delta t, x)-f(t, x)|<\varepsilon$$

于是当 Δt 充分小时，

$$E\{|f(t+\Delta t, Y(t+\Delta t))-f(t, Y(t+\Delta t))|^2\}$$
$$\leqslant E\{\varepsilon^2\}=\varepsilon^2$$

这就证明了 $f(t, Y(t))$ 的均方连续性.

现在用逐次逼近法来证明定理. 对 $t\in T$, 定义

$$X_0(t)=X_0$$

$$X_{n+1}(t)=X_0+\int_{t_0}^{t} f(s, X_n(s))ds$$
$$+\int_{t_0}^{t} g(s, X_n(s))dW(s), \qquad n\geqslant 0$$

首先要证明上面的定义是合理,即出现的积分都是存在的, 为此对 n 用归纳法来证明. 当 $n=0$ 时, $X_0(t)$ 是有定义的, 而且它显然具有如下两个性质:

(i) $X_0(t)$, $t\in T$ 是均方连续的;

(ii) 对 $t_0\leqslant s_1, s_2, s_3, s_4\leqslant t_{k-1}<t_k$ 有 $(X_0(s_1), X_0(s_2))$, $W(s_4)-W(s_3)$ 与 $W(t_k)-W(t_{k-1})$ 是独立的.

于是根据上面已证的结果, 由 $X_0(s)$ 是均方连续的, 可推知 $f(s, X_0(s))$ 是均方连续的,从而

$$Y_1(t)=\int_{t_0}^{t} f(s, X_0(s))ds$$

存在,并且易知它关于 t 是均方连续的,即 $Y_1(t)$ 具有性质(i). 由积分的定义,它是

$$\sum_j f(s_j, X_0(s_j))\Delta s_j$$

的均方极限, 诸 $X_0(s_j)$ 由于 $s_j\leqslant t<t_{k-1}<t_k$, 因此都与 $W(t_k)-W(t_{k-1})$ 独立, 从而上述和式, 乃至积分都与 $W(t_k)-W(t_{k-1})$ 独立,即 $Y_1(t)$ 具有性质(ii).

同样可知, $g(s, X_0(s))$ 是均方连续的, 并且具有性质(ii), 因

此根据定理 1，下列伊藤积分存在

$$Z_1(t)=\int_{t_0}^{t} g(s, X_0(s))dW(s)$$

并由定理 2 的系知，$Z_1(t)$ 是均方连续的，即具有性质(i)，完全同上面一样，可以证明 $Z_1(t)$ 具有性质(ii)．由于

$$X_1(t)=X_0+Y_1(t)+Z_1(t)$$

所以容易知道，$X_1(t)$ 也具有性质(i)，(ii)．由此同样地可以证明 $X_2(t), X_3(t), \cdots$ 都有定义，并且有性质(i)和(ii)．

下面来证明，上面定义的 $X_n(t)$ 对 $a\leqslant t\leqslant b$ 一致均方收敛于某个 $X(t)$，也就是要证明 $\{X_n(t)\}$ 对 $a\leqslant t\leqslant b$ 是一致均方基本序列．令

$$\Delta_n X(t)=X_{n+1}(t)-X_n(t)$$

$$\Delta_n f(t)=f(t, X_{n+1}(t))-f(t, X_n(t))$$

$$\Delta_n g(t)=g(t, X_{n+1}(t))-g(t, X_n(t))$$

于是，有

$$\Delta_0 X(t)=\int_{t_0}^{t} f(s, X_0(s))ds+\int_{t_0}^{t} g(s, X_0(s))dW(s)$$

$$\Delta_{n+1}X(t)=\int_{t_0}^{t} \Delta_n f(s)ds+\int_{t_0}^{t} \Delta_n g(s)dW(s), \qquad n\geqslant 0$$

由定理 2.8 的系，得

$$\begin{aligned}&\left\|\int_{t_0}^{t} f(s, X_0(s))ds\right\|^2\\ &\leqslant (t-t_0)\int_{t_0}^{t} \|f(s, X_0(s))\|^2 ds\\ &\leqslant (t-t_0)\int_{t_0}^{t} K^2(1+\|X_0(s)\|^2)ds\\ &=(t-t_0)K^2(1+\|X_0\|^2)(t-t_0)\\ &\leqslant A^2(t-t_0)\end{aligned}$$

其中 $A^2=K^2(1+\|X_0\|^2)(b-a)$，即

$$\left\|\int_{t_0}^{t} f(s, X_0(s))ds\right\| \leqslant A(t-t_0)^{\frac{1}{2}}$$

类似地，有

$$\left\|\int_{t_0}^{t} \Delta_n f(s)ds\right\|^2$$

$$\leqslant (t-t_0)\int_{t_0}^{t} \|\Delta_n f(s)\|^2 ds$$

$$\leqslant (b-a)\int_{t_0}^{t} K^2\|\Delta_n X(s)\|^2 ds$$

即

$$\left\|\int_{t_0}^{t} \Delta_n f(s)ds\right\|$$

$$\leqslant K\sqrt{b-a}\left(\int_{t_0}^{t} \|\Delta_n X(s)\|^2 ds\right)^{\frac{1}{2}}$$

由定理 1 的系，有

$$\left\|\int_{t_0}^{t} g(s, X_0(s))dW(s)\right\|^2$$

$$=\int_{t_0}^{t} \|g(s, X_0(s))\|^2 ds$$

$$\leqslant \int_{t_0}^{t} K^2(1+\|X_0(s)\|^2)ds$$

$$=K^2(1+\|X_0\|^2)(t-t_0)$$

于是有

$$\left\|\int_{t_0}^{t} g(s, X_0(s))dW(s)\right\| \leqslant B(t-t_0)^{\frac{1}{2}}$$

其中 $B=K\sqrt{1+\|X_0\|^2}$. 类似地，有

$$\left\|\int_{t_0}^{t} \Delta_n g(s)dW(s)\right\|^2$$

$$=\int_{t_0}^{t} \|\Delta_n g(s)\|^2 ds$$

$$\leqslant \int_{t_0}^{t} K^2\|\Delta_n X(s)\|^2 ds$$

即

$$\left\|\int_{t_0}^{t} \Delta_n g(s) dW(s)\right\| \leqslant K\left(\int_{t_0}^{t} \|\Delta_n X(s)\|^2 ds\right)^{\frac{1}{2}}$$

利用上述诸不等式，可得

$$\|\Delta_{n+1}X(t)\| \leqslant \left\|\int_{t_0}^{t} \Delta_n f(s) ds\right\| + \left\|\int_{t_0}^{t} \Delta_n g(s) dW(s)\right\|$$

$$\leqslant K(1+\sqrt{b-a})\left(\int_{t_0}^{t} \|\Delta_n X(s)\|^2 ds\right)^{\frac{1}{2}}, \quad n \geqslant 0$$

以及

$$\|\Delta_0 X(t)\| \leqslant A(t-t_0)^{\frac{1}{2}} + B(t-t_0)^{\frac{1}{2}}$$
$$= C(t-t_0)^{\frac{1}{2}}$$

于是用逐次迭代，可得:

$$\|\Delta_1 X(t)\|$$
$$\leqslant K(1+\sqrt{b-a})\left[\int_{t_0}^{t} \|\Delta_0 X(s)\|^2 ds\right]^{\frac{1}{2}}$$
$$\leqslant K(1+\sqrt{b-a})\left[\int_{t_0}^{t} C^2(s-t_0) ds\right]^{\frac{1}{2}}$$
$$= K\cdot(1+\sqrt{b-a}) C\left[\frac{(t-t_0)^2}{2!}\right]^{\frac{1}{2}}$$

一般地，有

$$\|\Delta_n X(t)\| \leqslant [K(1+\sqrt{b-a})]^n C\left[\frac{(t-t_0)^{n+1}}{(n+1)!}\right]^{\frac{1}{2}}$$

于是当 $m, n \to \infty$ 时有

$$\|X_n(t) - X_m(t)\| = \left\|\sum_{k=m}^{n-1} \Delta_k X(t)\right\|$$
$$\leqslant \sum_{k=m}^{n-1} \|\Delta_k X(t)\| \to 0$$

即 $\{X_n(t)\}$ 是均方基本序列，而且，当 t 在有限区间 $a \leqslant t \leqslant b$ 中时，关于 t 是一致的。因此，存在均方连续的 $X(t)$，使

$$\lim_{n\to\infty} X_n(t) = X(t)$$

收敛关于 t 是一致的, 再应用定理 3 即可证明存在性.

再来证明唯一性. 设 $X(t)$ 和 $Y(t)$ 是两个解, 记

$$\Delta(t) = X(t) - Y(t)$$

$$\Delta f(t) = f(t, X(t)) - f(t, Y(t))$$

$$\Delta g(t) = g(t, X(t)) - g(t, Y(t))$$

则有

$$\Delta(t) = \int_{t_0}^{t} \Delta f(s) ds + \int_{t_0}^{t} \Delta g(s) dW(s)$$

仿前可得:

$$\|\Delta(t)\| \leqslant K(1+\sqrt{b-a})\left(\int_{t_0}^{t} \|\Delta(s)\|^2 ds\right)^{\frac{1}{2}} \tag{4}$$

另外, 由于 $X(t)$ 和 $Y(t)$ 是均方连续的, 所以 $\|\Delta(s)\|$ 在 $a \leqslant s \leqslant b$ 上有界, 因而有

$$\int_{t_0}^{t} \|\Delta(s)\|^2 ds \leqslant D(t-t_0)$$

其中 $D = \max\limits_{a\leqslant t\leqslant b} \|\Delta(t)\|$, 于是

$$\|\Delta(t)\| \leqslant K(1+\sqrt{b-a}) D(t-t_0)^{\frac{1}{2}} \tag{5}$$

将(5)代入(4)右边, 得

$$\|\Delta(t)\| \leqslant K(1+\sqrt{b-a})\left(\int_{t_0}^{t} K^2(1+\sqrt{b-a})^2 D^2 (s-t_0) ds\right)^{\frac{1}{2}}$$

$$= K^2(1+\sqrt{b-a})^2 D\left(\int_{t_0}^{t} (s-t_0) ds\right)^{\frac{1}{2}}$$

$$= K^2(1+\sqrt{b-a})^2 D\left[\frac{(t-t_0)^2}{2!}\right]^{\frac{1}{2}}$$

再将此式代入(4)右边, 并反复进行, 可得

$$\|\Delta(t)\| \leqslant K^n(1+\sqrt{b-a})^n D\cdot\left[\frac{(t-t_0)^n}{n!}\right]^{\frac{1}{2}} \to 0, \qquad (n\to\infty)$$

唯一性证毕.

关于解的性质，我们有如下的

定理 5 在定理 4 的条件下，所得的唯一解 $X(t)$ 是马尔可夫过程.

［证明］ 我们要证明，对 $a\leqslant t_0<s<t\leqslant b$，$X(t)$在$X(s)$已知时，它与 $X(u)$，$a\leqslant u<s$ 无关. 事实上，由于积分的可加性，我们完全一样地可以证明

$$X(t)=X(s)+\int_s^t f(v,\ X(v))dv+\int_s^t g(v,X(v))dW(v)$$

因此 $X(t)$只依赖于 $X(s)$，$\{X(v),s\leqslant v\leqslant t\}$和$\{\Delta W(v),s\leqslant v\leqslant t\}$，其中 $\Delta W(v)=W(v+\Delta v)-W(v)$ 是 $W(v)$ 的增量，而 $\{X(v),s\leqslant v\leqslant t\}$ 也只是依赖于 $X(s)$ 和$\{\Delta W(v),\ s\leqslant v\leqslant t\}$的. 因此 $X(t)$，$t>s$ 只依赖于 $X(s)$ 和$\{\Delta W(v),s\leqslant v\leqslant t\}$. 同样，由

$$X(u)=X_0+\int_{t_0}^u f(v,X(v))dv+\int_{t_0}^u g(v,X(v))dW(v)$$

可知，$\{X(u),\ t_0\leqslant u\leqslant s\}$只依赖于 X_0 和$\{\Delta W(v),\ t_0\leqslant v\leqslant s\}$. 根据定理的假定$\{\Delta W(v),s\leqslant v\leqslant t\}$与 X_0，$\{\Delta W(v),\ t_0\leqslant v\leqslant s\}$是独立的. 因此，当 $X(s)$ 已知时，$X(t)$与$\{X(u),t_0\leqslant u\leqslant s\}$是独立的. 定理得证.

现在来看一个例子

$$mY'(t)=-f\cdot Y(t)+N(t),\qquad t\geqslant 0$$

其中 m，f 是正数，$N(t)$是正态白噪声. 这个方程有一个明显的物理解释. 假定 $Y(t)$ 表示时刻 t 时在液面上运动的一个粒子的某个方向上的速度，m 是它的质量，$-f\cdot Y(t)$表示由于粒子与液体摩擦而产生的阻尼力，$N(t)$表示由于分子冲击粒子所产生的总的合力，它当然是随机的，这里假定它是正态白噪声，这在物理上是比较符合实际情况的. 这样，$-fY(t)+N(t)$便是时刻 t 时作用在粒子上的总的合力，因而上述方程不过是牛顿第二定律的直

接应用.

上式也可改写成

$$dY(t)=-\frac{f}{m}Y(t)dt+dW(t)$$

其中 $W(t)$ 是维纳过程，即有

$$E\{W(t+s)-W(s)\}=0$$

$$E\{[W(t+s)-W(s)]^2\}=2Dt$$

$$E\{[W(t_2)-W(t_1)][W(s_2)-W(s_1)]\}=0$$

$$t_1<t_2\leqslant s_1<s_2$$

于是有

$$\Delta Y=Y(t+\Delta t)-Y(t)=\int_t^{t+\Delta t}dY(t)$$

$$=-\beta Y(t)\Delta t+W(t+\Delta t)-W(t)+o(\Delta t)$$

其中 $\beta=\frac{f}{m}>0$. 对上式两边取条件数学期望可得

$$E\{\Delta Y|Y(t)=y\}=-\beta y\Delta t+o(\Delta t)$$

和

$$E\{(\Delta Y)^2|Y(t)=y\}=\beta^2y^2\Delta t^2+2D\Delta t+o(\Delta t)$$

因此，有

$$\lim_{\Delta t\to 0}\frac{1}{\Delta t}E\{\Delta Y|Y(t)=y\}=-\beta y$$

$$\lim_{\Delta t\to 0}\frac{1}{\Delta t}E\{(\Delta Y)^2|Y(t)=y\}=2D$$

所以它是一个扩散过程，相应的偏移系数和扩散系数是

$$a(t,y)=-\beta y$$

$$b(t,y)=2D$$

于是由第二章§3第二段的例2立即可知，$Y(t)$ 是奥伦斯坦-乌伦贝克过程，其均值函数和协方差函数分别是

$$m(t)=0$$

$$\Gamma(s,\ t)=\frac{D}{\beta}e^{-\beta|t-s|}$$

如果用 $X(t)$ 表示时刻 t 时质点的位置,则有

$$X(t)-X(0)=\int_0^t Y(s)ds$$

由此可知 $X(t)$ 是正态过程,且有

$$E\{X(t)-X(0)\}=0$$

和

$$\begin{aligned}
&E\{|X(t)-X(0)|^2\}\\
&=\int_0^t\int_0^t \Gamma(u,v)dudv\\
&=\int_0^t\int_0^t \frac{D}{\beta}e^{-\beta|u-v|}dudv\\
&=2\frac{D}{\beta}\int_0^t\int_v^t e^{-\beta(u-v)}dudv\\
&=2\frac{D}{\beta^2}\int_0^t[1-e^{-\beta(t-v)}]dv\\
&=2\frac{D}{\beta^3}[\beta t-1+e^{-\beta t}]
\end{aligned}$$

习　题

1. 考虑如下定义的一族随机变量 $\{X_n,\ n\geq 1\}$,

$$X_n=\begin{cases} n, & 概率为\frac{1}{n^k}\\ 0, & 概率为\ 1-\frac{2}{n^k}\\ -n, & 概率为\frac{1}{n^k}\end{cases}$$

其中 k 是任一正数. 试分析 $\{X_n,\ n\geq 1\}$ 的收敛性:

(i) 对任意正的 k, X_n 以概率 1 收敛于 0;

(ii) 当 $k>1$ 时, X_n 均方收敛于 0;

(iii) 当 $k\leqslant 1$ 时，X_n 不均方收敛于 0.

2. **遍历性** 为简单起见，考虑均值为零的离散参数的随机过程$\{X_n, n\geqslant 1\}$. 再考虑样本平均序列:

$$M_n=\frac{1}{n}\sum_{i=1}^{n}X_i$$

则显然有 $E\{M_n\}=0$, 如果还有

$$\lim_{n\to\infty}E\{M_n^2\}=\lim_{n\to\infty}\mathrm{Var}\{M_n\}=0$$

则称样本平均序列是(均方)遍历的.

下面假定$\{X_n, n\geqslant 1\}$的协方差函数

$$\Gamma(m,n)=E\{X_m\cdot\overline{X}_n\}$$

是有界的，即存在常数 K, 使对一切 m, n 有 $|\Gamma(m,n)|\leqslant K$. 并令

$$C(n)=E\{X_n\cdot\overline{M}_n\}=\frac{1}{n}\sum_{i=1}^{n}E\{X_n\overline{X}_i\}$$

试证 (i) $|C(n)|^2\leqslant K\cdot\mathrm{Var}\{M_n\}$;

(ii) $n^2\cdot\mathrm{Var}\{M_n\}=2\sum_{i=1}^{n}i\cdot C(i)-\sum_{i=1}^{n}\mathrm{Var}\{X_i\}$;

(iii) 如 $\lim_{n\to\infty}C(n)=0$, 则

$$\lim_{n\to\infty}\frac{1}{n^2}\sum_{i=1}^{n}iC(i)=0$$

(iv) 利用上述结果，证明样本序列$\{M_n, n\geqslant 1\}$是遍历的充要条件是:

$$\lim_{n\to\infty}C(n)=0$$

3. 试研究下列随机过程的均方连续性，均方可微性和均方可积性:

(i) $X(t)=At+B$;

(ii) $X(t)=At^2+Bt+C$;

(iii) $X(t)$是普阿松过程;

(iv) $X(t)$是维纳-爱因斯坦过程;

(v) $X(t)$的均值为零，协方差函数为 $\Gamma(s,t)=e^{-a|s-t|}$, 其中 a 是正的常数;

(vi) $X(t)$的均值函数为零，协方差函数是

$$\Gamma(s,t)=\frac{1}{a^2+(s-t)^2}$$

4. 对上题中的随机过程，试求

$$Y(t)=\frac{1}{t}\int_0^t X(\tau)d\tau$$

和
$$Z(t)=\frac{1}{L}\int_t^{t+L} X(\tau)d\tau$$

的均值函数和协方差函数，其中 L 是一个正的常数.

5. 对第 3 题中的(i)，(ii)，(vi)三种过程，试求 $X'(t)$的均值函数和协方差函数.

6. 随机过程的切比雪夫不等式. 设$\{X(t), a\leqslant t\leqslant b\}$是实的均方可微的随机过程. 记

$$D(t)=\sqrt{E\{|X(t)|^2\}}$$
$$D_1(t)=\sqrt{E\{|X'(t)|^2\}}$$

试证：(i) $P\{\sup\limits_{a\leqslant t\leqslant b}|X(t)|>\varepsilon\}\leqslant\frac{1}{\varepsilon^2}E\{\sup\limits_{a\leqslant t\leqslant b}X^2(t)\}$;

(ii) $X^2(t)=X^2(a)+2\int_a^t X'(\tau)X(\tau)d\tau=X^2(b)-2\int_t^b X'(\tau)X(\tau)d\tau$;

(iii) $2X^2(t)\leqslant X^2(a)+X^2(b)+2\int_a^b|X'(\tau)X(\tau)|d\tau$ (注意右边与 t 无关!);

(iv) $$E\{\sup_{a\leqslant t\leqslant b}X^2(t)\}\leqslant\frac{E\{X^2(a)\}+E\{X^2(b)\}}{2}+\int_a^b\sqrt{E\{X'(\tau))^2\}E\{X^2(\tau)\}}d\tau$$

(v) 综上所述，可得下列随机过程的切比雪夫不等式:

$$P\{\sup_{a\leqslant t\leqslant b}|X(t)|>\varepsilon\}\leqslant\frac{D^2(b)+D^2(a)}{2}+\int_a^b D(\tau)D_1(\tau)d\tau$$

7. 设 Y_1 和 Y_2 是相互独立的均值为零方差为 1 的正态分布的随机变量. 定义二维随机变量

$$(X_1,X_2)=\begin{cases}(Y_1,|Y_2|), & \text{如 } Y_1\geqslant 0\\ (Y_1,-|Y_2|), & \text{如 } Y_1<0\end{cases}$$

试证：(i) X_1 和 X_2 都是正态分布的;

(ii) (X_1, X_2)不是二维正态分布.

8. 设(X_1, X_2, X_3, X_4)是均值为零的四维正态分布随机变量，试证：

$$E\{X_1X_2X_3X_4\}=E\{X_1X_2\}E\{X_3X_4\}+E\{X_1X_3\}E\{X_2X_4\}$$
$$-E\{X_1X_4\}E\{X_2X_3\}$$

9. 设$(X(t), Y(t))$是零均值的二维正态过程；试证：

(i) $\alpha X(t)+\beta Y(t)$也是正态过程，其中α, β是两个常数. 并写出它的协方差函数；

(ii) $X^2(t)$是二阶矩过程，并写出它的均值函数和协方差函数.

10. 考虑伊藤积分

$$(I)\int_a^b X(t)dW(t)$$

试证，如果$X(t)$退化为一个决定性的连续函数，则伊藤积分就变为黎曼-司蒂吉斯均方积分.

11. 试计算下列期望：

$$E\left\{(I)\int_a^b X(t)dW(t)\right\}$$

$$E\left\{[(I)\int_a^b X(t)dW(t)][(I)\int_a^b Y(t)dW(t)]\right\}$$

其中$\{X(t), t\in T\}$和$\{Y(t), t\in T\}$是$T=[a, b]$上均方连续随机过程，而且$X(t)$和$Y(t)$是满足定理 4.1 条件的.

12. 试证定理 2.16.

第四章 平稳过程

在上一章，我们主要从随机过程的均值和协方差函数这两个数量特征上来规定它的性质，并讨论它的统计规律性．在这一章，我们也将用类似的方法来讨论更为特殊的一类二阶矩过程——平稳过程，为了以后讨论方便，我们先在§1把第一章、第三章中提到过的有关平稳过程的内容回顾一下，并作一些必要的补充．

§1 平稳过程和协方差函数

一、平稳过程的定义

在自然科学和工程技术中，我们常常会遇到一类与前面所述的马尔可夫过程不一样的随机过程，从随机过程本身随时间的变化和相互联系来看，不仅它的当时情况而且它的过去情况都对未来有着不可忽视的影响．例如在纺纱过程中，我们可将棉纱的截面积看作截点坐标的随机过程．由于棉纱是一束束纤维组成的，因此在点 x_1 处的截面积的大小影响到 x_2（大于 x_1）处的截面积大小，且当 x_2 愈接近 x_1 时，影响愈大．再如飞机沿着航线在某一高度飞行时，由于受到湍流的影响而产生的随机波动．因为飞机飞行时有很大的惯性，所以我们不能把这个波动过程看为马尔可夫过程．从随机过程前后联系上来研究随机过程本身变化的统计规律时，很重要的一类就是平稳过程．所谓平稳过程粗略地说就是指它是统计特征不随时间推移而变化的随机过程．

如同第一章所说的，对于随机过程，也象随机变量一样，我们

要完整地了解它的有限维分布函数族的全部知识是困难的，而了解它的某些数字特征却是可能的. 另一方面，就利用这些数字特征的知识，也可能解决实用中所关心的随机过程有关的各种问题.

定义 1　设复值随机过程 $X_T=\{X(t), t\in T\}$ 存在有限二阶矩 $E|X(t)|^2<\infty$, T 为实数集或其子集. 若 X_T 满足

$$EX(t)=m(\text{常数}) \tag{1}$$

$$E\{X(t)\overline{X(s)}\}=\Gamma(t-s) \tag{2}$$

则称 X_T 为**平稳随机过程**，简称**平稳过程**. 当 T 为整数集或 $\{n\Delta t, n=0, 1, \cdots\}$ 时，也称 X_T 为**平稳时间序列**.

$$\begin{aligned} B(t-s)&=E(X(t)-m)(\overline{X(s)-m}) \\ &=\Gamma(t-s)-m^2 \end{aligned} \tag{3}$$

也称为 X_T 的**协方差函数**.

(1) 表明 $X(t)$ 的均值是不随时间而改变的，即均值函数取常数值. 不仅如此，(2)表明 $(X(t), X(s))$ 间的二阶统计特征与 $(X(t+h), X(s+h))$ 间相应的统计特征是相同的，即不随时间推移而改变. 当然还可以要求 X_T 的高阶矩甚至一切有限维分布随时间推移是不变的，这是比定义 1 更强的平稳性. 为了区别于这种强的平稳性要求，定义 1 规定的平稳性也称为**宽平稳**、**弱平稳**或**广义平稳**的. 但在本书中主要只涉及这种常用的较弱的平稳性，所以平稳过程这一名词就是指满足定义 1 规定的宽平稳过程.

此外，在工程技术上也常把(3)中的 $B(t)$ 称为相关函数，但对照随机变量间协方差与相关这两个名词的用法，这里仍把(3)中 $B(t)$ 称为 X_T 的协方差函数，而把

$$R(t)=\frac{B(t)}{B(0)}$$

称为**标准化相关函数**.

在自动控制和电子学中分析电路中的电压或电流变化时，由

于自由电子的不规则运动(即热运动)会引起电路中的电压或电流的随机波动. 物理学中也称这种波动为热扰动. 它的统计特性可视为不随时间推移而变化的, 因此这一波动过程就是平稳随机过程, 它也是电路中的一种噪声.

在考虑船舶受海浪冲击而产生摇摆和其它适航性问题中, 就要把海浪看为随机过程. 一般说在某个地理位置上海浪的波高就是一个平稳随机过程, 它的统计特征不随时间推移而改变, 不仅如此, 若把同一时间海浪波高 h 看为位置坐标 (x, y) 的函数, $h(x, y)$ 也是随机函数, 由于它依赖于两个独立坐标 x, y, 所以也称为随机场, 在一定范围内, $h(x, y)$ 的统计特性也不随地理位置的推移而改变, 所以 $h(x, y)$ 按空间来说是平稳的, 我们常用齐次这一名词来形容它, 所以海浪的波高 $h(x, y)$ 是一个齐次随机场. 要是讨论大气湍流的统计特性时, 还会涉及三个变量的齐次随机场, 虽然它们比一般的平稳过程更复杂, 但处理平稳过程的一些方法大多也能适用于齐次随机场.

下面我们再举一些简单的平稳过程的例子.

[例 1] 若 $X=\{X(n), n=0, \pm 1, \cdots\}$ 为实的互不相关随机变量序列, 且 $EX(n)=0, DX(n)=\sigma^2$, 则因为

$$EX(n)=0=\text{常数}$$

$$EX(n)X(m)=\begin{cases}\sigma^2, & n=m\\ 0, & n\neq m\end{cases}$$

即协方差函数只与 $n-m$ 有关, 所以它是一个平稳时间序列.

在物理和工程技术上, 常把上述随机变量序列称为"白噪声". 它普遍地存在于各种波动现象中(如电流的波动, 电子发射的波动和通信设备各部分电流或电压的波动等), 又具有较为简单的数学模型, 所以常被作为一种典型的噪声模型来考虑.

[例 2] 设 $X(n)$ 是例 1 中的随机变量序列, $\{C_n\}$ 为满足

$$\sum_{n=-\infty}^{\infty}|C_n|^2<\infty \tag{4}$$

的复数序列，则

$$Y(n)=\sum_{k=-\infty}^{\infty}C_kX(n-k) \tag{5}$$

是一个平稳时间序列. 首先，我们先来说明，对每个 n，(5)是均方收敛的，若记

$$Y^{(N)}(n)=\sum_{k=-N}^{N}C_kX(n-k)$$

则当 $N>M$ 时，

$$\begin{aligned}
&E|Y^{(N)}(n)-Y^{(M)}(n)|^2\\
&=E|\sum_{M<|k|\leqslant N}C_kX(n-k)|^2\\
&=\sigma^2\sum_{M<|k|\leqslant N}|C_k|^2
\end{aligned}$$

因此由(4)及定理 3.1.6 可知，必存在随机变量 $Y(n)$，使 $Y^{(N)}(n)$ 均方收敛于 $Y(n)$，即随机变量序列的级数(5)是均方收敛的，所以用(5)来规定 $Y(n)$ 是合理的.

其次，利用均方收敛随机变量序列的性质(定理 3.1.7)，可得

$$\begin{aligned}
EY(n)&=\lim_{N\to\infty}E\{Y^{(N)}(n)\}=\lim_{N\to\infty}\sum_{|k|\leqslant N}C_kE\{X(n-k)\}\\
&=\sum_{k=-\infty}^{\infty}C_kE\{X(n-k)\}=0
\end{aligned}$$

$$\begin{aligned}
E\{Y(n)\overline{Y(m)}\}&=E\left(\sum_{k=-\infty}^{\infty}C_kX(n-k)\right)\overline{\left(\sum_{l=-\infty}^{\infty}C_lX(m-l)\right)}\\
&=\sum_{k=-\infty}^{\infty}\sum_{l=-\infty}^{\infty}C_k\overline{C_l}EX(n-k)\overline{X(m-l)}\\
&=\sum_{k=-\infty}^{\infty}C_k\overline{C}_{m-n+k}\sigma^2
\end{aligned}$$

$$=\sigma^2\left(\sum_{k=-\infty}^{\infty}C_{n-m+k}\overline{C_k}\right)$$

$$=B(n-m)$$

即均值为常数零，协方差函数只与 $n-m$ 有关，所以，$Y_T=\{Y(n)\}$ 为一平稳时间序列．形为(5)的 $Y(n)$ 又称为**滑动和**．

［例 3］ 与例 2 相类似的连续参数情形是当 $X_T=\{X(t), -\infty<t<\infty\}$ 是均值为零的正交增量过程，且

$$E|X(t)-X(s)|^2=\sigma^2(t-s)$$

又 $C(t)$ 为复值函数，且满足

$$\int_{-\infty}^{\infty}|C(t)|^2dt<\infty$$

则同例 2 一样，可以说明

$$Y(t)=\int_{-\infty}^{\infty}C(t-s)dX(s) \tag{6}$$

也是一个平稳过程，且

$$EY(t)=0$$

$$E\{Y(t)\overline{Y(s)}\}=\sigma^2\int_{-\infty}^{\infty}C(t-s+\tau)\overline{C(\tau)}d\tau$$

对于连续参数情形，形为(6)的过程也称为**滑动和**．

［例 4］ 若 $\{U_n, 1\leqslant n\leqslant N\}$，$\{V_n, 1\leqslant n\leqslant N\}$ 为互不相关的实随机变量序列

$$\begin{aligned}&EU_n=EV_n=0\\&EU_nV_m=0,\quad 1\leqslant m,\ n\leqslant N\\&EU_nV_m=EV_nV_m=\sigma_n^2\delta_{nm}\end{aligned} \tag{7}$$

又 $\omega_1, \omega_2, \cdots, \omega_N$ 为任意正实数，令

$$X(t)=\sum_{n=1}^{N}(U_n\cos(\omega_n t)+V_n\sin(\omega_n t)),\quad -\infty<t<\infty \tag{8}$$

这时

$$EX(t)=0$$

$$EX(t)X(s)$$

$$=\sum_{n=1}^{N}\sigma_n^2[\cos(\omega_n t)\cos(\omega_n s)+\sin(\omega_n t)\sin(\omega_n s)]$$

$$=\sum_{n=1}^{N}\sigma_n^2\cos\omega_n(t-s) \tag{9}$$

所以它是一个平稳过程.

［例 5］ 若$\{Z_n, n=0, \pm1, \cdots\}$为复的随机变量序列

$$EZ_n=0$$

$$EZ_n\overline{Z_m}=\sigma_n^2\delta_{nm} \tag{10}$$

且

$$\sum_{n=-\infty}^{\infty}\sigma_n^2<\infty$$

则同例 2 一样，对每个 t，下列级数是在均方意义下收敛的，

$$\sum_{n=-\infty}^{\infty}Z_n e^{i\omega_n t}$$

其中$\{\omega_n\}$为任意的一列实数. 因此我们可以规定

$$X(t)=\sum_{n=-\infty}^{\infty}Z_n e^{i\omega_n t} \tag{11}$$

利用随机变量级数均方收敛的性质，不难推出

$$EX(t)=0$$

$$EX(t)\overline{X(s)}=\sum_{n=-\infty}^{\infty}\sigma_n^2 e^{i\omega_n(t-s)} \tag{12}$$

所以 $X_T=\{X(t), -\infty<t<\infty\}$也是平稳过程.

上述两个例子对我们是十分重要的，因为 $\cos\omega t$， $\sin\omega t$， 或 $e^{i\omega t}$ 都是描述简谐振动的，$U_n\cos\omega_n t$，$V_n\sin\omega_n t$ 或 $Z_n e^{i\omega_n t}$ 都可以看为一个具有随机振幅的简谐振动. 上述例子表明，若不同频率

振动的振幅互不相关，那么这种简谐振动的有限项甚至无限项的迭加(只要它是均方收敛的)都是平稳过程. (9)和(12)又表明，它们的协方差函数亦有类似的分解，且可表为与随机振动有相同频率成分的简谐振动之和(即(11)、(12)有相同的 ω_n). 关于平稳过程和协方差函数类似的表示，在下一节我们还将作进一步的讨论.

[例 6] 随机电报信号. 在电报信号传输中，信号是由不同的电流符号给出，电流的发送又有一任意的持续时间，若电路中的电流 $X(t)$ 的变化可用图 1 来表示.

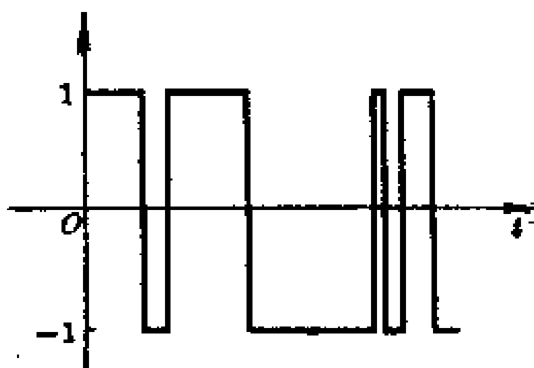


图 1

这种电流变化的特征是:

(i) 电流值只取 $+1$ 或 -1;

(ii) 电流符号变换的时间是随机的，在任一个时间区间 $(t, t+h)$ 内发生变号这一事件流是满足简单随机事件流的条件，由第一册 §2.4 的结论，在 $(t, t+h)$ 内出现 k 次变号的概率是服从普阿松分布的，

$$P(k)=\frac{(\mu h)^k}{k!}e^{-\mu h}$$

则当在 $(t, t+h)$ 中出现偶数次变号，$X(t)$ 和 $X(t+h)$ 取同号，在 $(t, t+h)$ 中出现奇数次变号，$X(t)$ 和 $X(t+h)$ 取异号，又若 ξ 为随机变量，它与 $X(t)$ 相互独立，

$$P(\xi=+a)=P(\xi=-a)=\frac{1}{2}$$

则 $E\xi=0, E\xi^2=a^2$，若我们考虑

$$Y(t)=\xi X(t)$$

则由于 ξ 与 $X(t)$ 相互独立，

$$E\{Y(t)\}=E\{\xi\}E\{X(t)\}=0 \tag{13}$$

$$E\{Y(t)Y(s)\}=E\{\xi^2\}E\{X(t)X(s)\}$$

$$=a^2\left\{\sum_{k=0}^{\infty}\frac{[\mu(t-s)]^{2k}}{(2k)!}e^{-\mu(t-s)}-\sum_{k=0}^{\infty}\frac{[\mu(t-s)]^{2k+1}}{(2k+1)!}e^{-\mu(t-s)}\right\}$$

$$=a^2e^{-2\mu(t-s)},\qquad t>s \tag{14}$$

所以 $Y_T=\{Y(t),-\infty<t<\infty\}$ 是平稳过程.

二、平稳过程的简单性质

从平稳过程的定义容易看出，它是一个二阶矩过程，所以第一、三章中对二阶矩过程得到的性质和结论对于平稳过程也同样成立. 不仅如此，有时还可得到更强的结论. 下面我们把 §3.1中二阶矩过程的一些结论转述到平稳过程上来.

性质 1 若 $B(\tau)$ 是平稳过程 $X_T=\{X(t),\ t\in T\}$ 的协方差函数，则

$$B(0)\geqslant 0$$

$$B(-\tau)=\overline{B(\tau)}$$

$$|B(\tau)|\leqslant B(0)$$

更一般的，$B(\tau)$ 是一个非负定函数，即对任何正整数 n，复数 $a_1,a_2,\cdots,a_n$ 及 T 中的 $t_1,t_2,\cdots,t_n$，都有不等式

$$\sum_{i,j=1}^{n}B(t_i-t_j)a_i\overline{a_j}\geqslant 0 \tag{15}$$

性质 2 若 $B(\tau)$ 是平稳过程 $X_T=\{X(t),\ -\infty<t<\infty\}$ 的协方差函数，则 X_T 均方连续的充要条件是 $B(\tau)$ 在 $\tau=0$ 连续，且这时 $B(\tau)$ 处处连续.

性质 3 若 $B(\tau)$ 是平稳过程 $X_T=\{X(t),\ -\infty<t<\infty\}$ 的协方差函数，则 X_T 为 p 次均方可微的充要条件是 $B(\tau)$ 在 $\tau=0$ 时 $2p$ 次可微，且这时 $B(\tau)$ 处处 $2p$ 次可微，并成立

$$E\{X^{(q)}(t)\overline{X^{(r)}(s)}\}=(-1)^r B^{(q+r)}(t-s)$$

$$0\leqslant q,\ r\leqslant p \tag{16}$$

性质 4 若 $X_T=\{X(t),\ -\infty<t<\infty\}$为均方连续平稳过程, $f(t)$为分段连续函数, 则在任何有限区间上, 下列积分在均方意义下存在:

$$\int_a^b f(t)X(t)dt$$

且对任一分段连续函数 $g(t)$ 和存在方差的随机变量Y有

$$\operatorname{cov}\left(\int_a^b f(t)X(t)dt,\ \int_a^b g(s)X(s)ds\right)$$

$$=\int_a^b\int_a^b B(t-s)f(t)\overline{g(s)}dsdt \tag{17}$$

$$E\left\{\int_a^b f(t)X(t)dt\cdot\overline{Y}\right\}$$

$$=\int_a^b f(t)E\{X(t)\overline{Y}\}dt \tag{18}$$

性质 5 设 X_T 为均方可微的实平稳过程, 则

$$E\{X(t)\cdot X'(t)\}=0$$

[证明] 见 § 3.3 第二段的引理.

三、平稳正态过程

在第一章中曾提到, 一个二元函数 $\Gamma(t,s)$ 可作为二阶矩过程协方差函数的充要条件是 $\Gamma(t,s)$为非负定的, 并对任一非负定函数可作出以此非负定函数为协方差函数的正态过程. 上面又提到平稳过程的协方差函数 $B(t)$ 是一元非负定函数. 若记

$$\Gamma(t,s)=B(t-s)$$

则由(15), 二元函数 $\Gamma(t,s)$为非负定的, 所以必存在正态过程 $X_T=\{X(t),t\in T\}$, 它的均值为零, 且以 $B(t-s)$为它的协方差函数, 这就表明一元函数$B(t)$ 可作为平稳过程的协方差函数的充要条件是 $B(t)$ 为非负定函数, 即满足(15).

若 X_T 为实平稳过程，且

$$E\{X(t)\}=m \tag{19}$$

$$E\{X(t)X(s)\}=B(t-s)+m^2$$

它又是正态过程，即它的有限维分布 $F(t_1,\cdots,t_n;x_1,\cdots,x_n)$ 都是正态分布，这时我们可写出 $F(t_1,\cdots,t_n;x_1,\cdots,x_n)$ 所对应的特征函数为：

$$\varphi(t_1,\cdots,t_n;u_1,\cdots,u_n)$$

$$=\exp\left\{i\left(\sum_{j=1}^{n}u_j\right)m-\frac{1}{2}\sum_{j,k=1}^{n}u_ju_kB(t_j-t_k)\right\} \tag{20}$$

从(20)不难看出，对任一 h

$$\varphi(t_1+h,\cdots,t_n+h;u_1,\cdots,u_n)=\varphi(t_1,\cdots,t_n;u_1,\cdots,u_n)$$

又因分布函数是由其特征函数唯一确定，必有

$$F(t_1+h,\cdots,t_n+h;x_1,\cdots,x_n)=F(t_1,\cdots,t_n;x_1,\cdots,x_n) \tag{21}$$

这就表明 X_T 的一切有限维分布都不随时间推移而改变，这是较之于(1)，(2)更强的一种平稳性，通常也称为**严平稳**，**狭义平稳**或**强平稳**的．上述说明表明对于正态过程来说，(1)，(2)规定的平稳性和(21)规定的平稳性是一致的．

当我们讨论平稳过程有关的各种问题时，我们只是讨论满足(1)，(2)规定的随机过程的种种特性，即使具体给出了一个满足(1)，(2)的均值和协方差函数，往往也不是只给出了一个平稳过程，而是一类具有相同的二阶统计特征的随机过程，所讨论的也是这一类过程所共有的特性，而上面的说明表示在这类过程中，总有一个正态过程．在平稳过程的研究中，就往往把这个平稳正态过程作为重要的特例，这时由于正态过程由其二阶特征可唯一确定其分布，因此从平稳正态过程可以得到较一般平稳过程更深刻的

结论，它往往还是深入研究平稳过程的出发点。

四、平稳正态马尔可夫过程

在§3.3曾提到正态马尔可夫过程，其标准化的相关函数必满足

$$R(r,t)=R(r,s)R(s,t),\quad r\leqslant s\leqslant t$$

若这一过程还是平稳的，则其标准化相关函数 $R(t)$ 必有

$$R(t+s)=R(t)R(s),\quad t,s>0 \tag{22}$$

对上式的进一步分析便可得到关于平稳正态马尔可夫过程的下列定理。

定理 1 (i) 若 $X_T=\{X(n),\ n=0,\pm1,\cdots\}$ 为平稳正态时间序列，$B(0)\neq 0$，则它是马尔可夫过程的充要条件是其协方差函数 $B(n)$ 可表为

$$B(n)=a^nB(0),\qquad n\geqslant 0,\ |a|\leqslant 1 \tag{23}$$

(ii) 若 $X_T=\{X(t),\ -\infty<t<\infty\}$ 为平稳正态过程，则它是均方连续的马尔可夫过程的充要条件是其协方差函数 $B(t)$ 可表为

$$B(t)=e^{at}B(0),\ t\geqslant 0,\ \mathrm{Re}\,a\leqslant 0 \tag{24}$$

［证明］ (i) 由(22)不难推出，若 X_T 为马尔可夫过程，必有

$$\frac{B(n)}{B(0)}=\left(\frac{B(1)}{B(0)}\right)^n$$

令 $a=\dfrac{B(1)}{B(0)}$，即得(23)。反之，由(23)不难推出 X_T 的标准化相关函数 $R(n)=\dfrac{B(n)}{B(0)}=a^{|n|}$ 满足(22)。

(ii) 若 X_T 为平稳正态马尔可夫过程，则其标准化相关函数 $R(t)$ 满足(22)，又因为 X_T 是均方连续的，所以 $R(t)$ 是 t 的连续函数。由此可知(见第一册§2.4引理)，

$$R(t)=e^{at},\quad t>0$$

又由标准化相关函数的性质可知$|R(t)|\leqslant 1$,故a满足

$$\mathrm{Re}\, a\leqslant 0$$

反之，当$R(t)$满足(24)时容易验证(22)也成立，即X_T为马尔可夫过程.又因(24)中的$R(t)$在$t=0$连续，所以X_T还是均方连续的．定理证毕.

关于平稳正态马尔可夫过程除了它的协方差函数满足上述条件外,其本身亦有特殊的表示.

定理 2 (i) 若$X_T=\{X(n),\ n=0,\pm 1,\cdots\}$是均值为零的平稳过程,则它的协方差函数$B(n)$可表为

$$B(n)=a^n B(0),\ n>0,\ |a|<1 \tag{25}$$

的充要条件是$X(n)$满足下列一阶差分方程

$$X(n)-aX(n-1)=\xi(n),\quad |a|<1 \tag{26}$$

其中$\xi(n)$为互不相关的平稳序列,且满足

$$E\{\xi(n)\overline{\xi(m)}\}=\sigma^2\delta_{nm} \tag{27}$$

或者等价地$X(n)$可表为

$$X(n)=\sum_{k=0}^{\infty}a^k\xi(n-k) \tag{28}$$

其中$\xi(n)$亦满足(27).

(ii) 若上述X_T还是正态过程,则(26)——(28)中的$\{\xi(n)\}$为独立同分布的$N(0,\sigma^2)$随机变量序列.

[证明] (i) 若X_T的协方差函数$B(n)$满足(25),令

$$\xi(n)=X(n)-aX(n-1)$$

则当$n\geqslant m$时,

$$\begin{aligned}E\xi(n)\overline{\xi(m)}&=E(X(n)-aX(n-1))\overline{(X(m)-aX(m-1))}\\&=B(n-m)-aB(n-m-1)-\bar{a}B(n-m+1)\\&\qquad+|a|^2B(n-m)\\&=\begin{cases}(1-|a|^2)B(0), & n=m\\ 0, & n>m\end{cases}\end{aligned}$$

所以$\{\xi(n)\}$满足(27), $X(n)$满足一阶差分方程(26)。

若 X_T 满足(26), (27)则

$$X(n)-aX(n-1)=\xi(n)$$
$$X(n-1)-aX(n-2)=\xi(n-1)$$
$$\cdots\cdots$$
$$X(n-M)-aX(n-M-1)=\xi(n-M)$$

即

$$X(n)=\sum_{k=0}^{M}a^k\xi(n-k)+a^{M+1}X(n-M-1)$$

由于$|a|<1$, $\sum_{k=0}^{\infty}|a|^{2k}<\infty$, 所以(28)中的级数是均方收敛的. 又

$$\lim_{M\to\infty}E|a^{M+1}X(n-M-1)|^2=\lim_{M\to\infty}a^{2(M+1)}E|X(0)|^2=0$$

所以

$$X(n)=\sum_{k=0}^{\infty}a^k\xi(n-k)$$

即 $X(n)$可表为(28).

反之, 若(27), (28)成立时, 当 $n\geqslant 0$,

$$B(n)=E\{X(n)\overline{X(0)}\}$$
$$=E\left\{\left(\sum_{k=0}^{\infty}a^k\xi(n-k)\right)\overline{\left(\sum_{l=0}^{\infty}a^l\xi(-l)\right)}\right\}$$
$$=\sum_{k,l=0}^{\infty}a^k\bar{a}^lE(\xi(n-k)\overline{\xi(-l)})$$
$$=\sum_{l=0}^{\infty}a^{n+l}\bar{a}^l$$
$$=a^n\sum_{l=0}^{\infty}|a|^{2l}=\frac{a^n}{1-|a|^2}$$

特别当 $n=0$ 时,

$$B(0)=\frac{1}{1-|a|^2}$$

所以，当 $n\geqslant 0$ 时，

$$B(n)=a^n B(0)$$

即 X_T 的协方差函数满足(25)。

(ii) 是正态分布随机变量独立与不相关等价这一事实的直接推论。

上述定理是叙述参数离散时平稳正态马尔可夫过程的表示，对于连续参数亦有类似的结论，即 X_T 的协方差函数满足(24)的充要条件是 $X(t)$满足下列一阶线性微分方程

$$X(t)-aX'(t)=\eta'(t) \tag{29}$$

或等价地 $X(t)$可表为下列滑动和

$$X(t)=\int_{-\infty}^{t} e^{a(t-s)}d\eta(s)$$

其中$\{\eta(s),-\infty<s<\infty\}$为满足下列条件的不相关增量过程

$$D(\eta(t)-\eta(s))=\sigma^2|t-s|$$

对于(29)中的 $\eta'(t)$需要一些特殊的说明，粗糙地说它是不相关增量过程的"导数"。特别当 $X(t)$是正态过程时，$\eta(t)$就是第二章中提到的维纳过程，$\eta'(t)$就是白噪声，而(29)中的微分方程，就可按§3.4中的随机微分方程来理解，而此时的 $X(t)$ 就是§2.3中提到的爱伦菲斯特-乌伦贝克过程。

§2. 平稳过程和协方差函数的谱分解

在富利埃分析中已从数学上论证了这样的事实，任一表示位移的时间的函数(周期的或非周期的)都可看为无数个(有限或无限个)简谐振动的迭加，对于描述随机变化的随机过程是否也有相类似的结论呢。§1的例4、例5已表明具有互不相关随机振幅简

谐振动的迭加一般是一个平稳过程，那么对平稳过程是否也有类似于富利埃分析中的分解定理呢，即对任一个平稳过程是否都可分解为随机的简谐振动的迭加呢，这就是本节中平稳过程谱分解定理所回答的问题. 首先我们要介绍协方差函数的谱分解定理.

定理 1　若 $B(n)$ 为平稳时间序列 $X_T=\{X(n), n=0, \pm 1, \cdots\}$ 的协方差函数，则它可表示为

$$B(n)=\int_{-\pi}^{\pi} e^{in\lambda}dF(\lambda) \tag{1}$$

其中 $F(\lambda)$ 是左连续，有界，非降函数.

特别对实平稳序列的协方差函数 $B(n)$ 必可表为

$$B(n)=\int_{0}^{\pi} \cos n\lambda dF_1(\lambda) \tag{2}$$

且(1)，(2)中的 $F(\lambda)$，$F_1(\lambda)$ 不计常数之差都由 $B(n)$ 唯一确定.

［证明］　若 $B(n)$ 为 X_T 的协方差函数，则由(1.15)，它是一个正定序列，由赫尔格洛兹定理(第一册定理 5.2.10)可知，必存在左连续非降有界函数 $F(\lambda)$，使 $B(n)$ 可表为

$$B(n)=\int_{-\pi}^{\pi} e^{in\lambda}dF(\lambda)$$

且不计常数之差 $F(\lambda)$ 由 $B(n)$ 唯一确定.

若 X_T 为实的，则 $B(n)$ 也是实的，所以 $B(n)=B(-n)$，而

$$\int_{-\pi}^{\pi} e^{in\lambda}dF(\lambda)=B(n)=B(-n)$$

$$=\int_{-\pi}^{\pi} e^{-in\lambda}dF(\lambda)$$

所以

$$F(\lambda_2)-F(\lambda_1)=F(-\lambda_1+0)-F(-\lambda_2+0)$$

因此

$$B(n)=\frac{1}{2}\left(\int_{-\pi}^{\pi} e^{in\lambda}dF(\lambda)+\int_{-\pi}^{\pi} e^{-in\lambda}dF(\lambda)\right)$$

$$=\int_{-\pi}^{\pi}\cos n\lambda dF(\lambda)$$

$$=2\int_{0}^{\pi}\cos n\lambda dF(\lambda)$$

取 $F_1(\lambda)=2F(\lambda)+C$ 即得(2)，$F_1(\lambda)$ 的唯一性亦可由(1)中 $F(\lambda)$ 的唯一性推出.

对于连续参数情形有如下定理.

定理 2 若 $B(t)$为均方连续平稳过程 X_T 的协方差函数，则它可表示为

$$B(t)=\int_{-\infty}^{\infty}e^{it\lambda}dF(\lambda) \tag{3}$$

其中 $F(\lambda)$是左连续有界非降函数.

特别对实平稳过程的协方差函数 $B(t)$必可表为

$$B(t)=\int_{0}^{\infty}\cos t\lambda dF_1(\lambda) \tag{4}$$

且(3)，(4)中的 $F(\lambda)$和 $F_1(\lambda)$不计常数之差都由 $B(t)$唯一确定.

［证明］ 若 $B(t)$为均方连续 X_T 的协方差函数，则 $B(t)$是连续正定函数，由波赫纳尔-辛钦定理(第一册定理 5.2.9)它必可表为(3)的形式，且 $F(\lambda)$由 $B(t)$唯一确定(不计常数之差)，至于实的情况也可象定理 1 一样证明.

定理 1 和定理 2 又称为协方差函数的谱分解定理，其中的 $F(\lambda)$或 $F_1(\lambda)$都称为平稳过程 X_T 的**谱函数**. 特别当 $F(\lambda)$或 $F_1(\lambda)$绝对连续时，即它可表为

$$F(\lambda)=\int_{-\infty}^{\lambda}f(\mu)d\mu+C$$

则称 $f(\lambda)$为 X_T 的**谱密度**，且此时协方差函数的谱分解式(3)可写为

$$B(t)=\int_{-\infty}^{\infty}e^{i\lambda t}f(\lambda)d\lambda \tag{5}$$

离散参数和实值过程也有类似的表示.

对于平稳过程而言,除了常数均值外,协方差函数就是它的最重要的统计特征了,而定理1,2又表明协方差函数和谱函数是互相单值确定的. 因此给定了过程的均值和谱函数也就完整地给出了过程的统计特征,因而关于平稳过程的许多讨论,都要把问题的条件和结论与其谱函数联系起来. 今后我们就常常这样做.

在利用协方差函数确定谱函数时一个重要的特殊情况是

$$\int_{-\infty}^{\infty}|B(\tau)|d\tau<\infty$$

这时必存在连续谱密度 $f(\lambda)$(可参见第一册§4.4相仿的定理),且有

$$f(\lambda)=\frac{1}{2\pi}\int_{-\infty}^{\infty}e^{-i\lambda t}B(t)dt \tag{6}$$

在实的情况下,

$$f_1(\lambda)=\frac{2}{\pi}\int_0^{\infty}\cos\lambda t B(t)dt,\ 0\leqslant\lambda<\infty$$

对离散参数情况当 $\sum\limits_{n=-\infty}^{\infty}|B(n)|<\infty$ 时亦有类似的结论,即

$$f(\lambda)=\frac{1}{2\pi}\sum_{n=-\infty}^{\infty}B(n)e^{-in\lambda} \tag{7}$$

在实的情况下,

$$f_1(\lambda)=\frac{1}{\pi}\sum_{n=0}^{\infty}B(n)\cos n\lambda$$

[例1] 若 $X=\{X(n),n=0,\pm1,\cdots\}$ 为零均值互不相关随机变量序列,则

$$B(n)=\begin{cases}\sigma^2, & n=0\\ 0, & n\neq0\end{cases}$$

因为 $\sum\limits_{n}|B(n)|<\infty$,利用(7)可知它存在谱密度 $f(\lambda)$,且

$$f(\lambda)=\frac{1}{2\pi}\sum_{n=-\infty}^{\infty}B(n)e^{-in\lambda}$$

$$=\frac{\sigma^2}{2\pi},\ -\pi\leqslant\lambda\leqslant\pi \tag{8}$$

而谱函数可取为 $F(\lambda)=\frac{\sigma^2}{2\pi}(\lambda+\pi)$.

[例 2] 若$\{Y(n)\}$由§1 例 2 规定, $\{X(n)\}$ 是例 1 中的随机变量序列, 且

$$\sum_{n=-\infty}^{\infty}|C_n|<\infty$$

$$Y(n)=\sum_{k=-\infty}^{\infty}C_{n-k}X(k) \tag{9}$$

则 $Y(n)$的协方差函数为

$$B(n)=\sigma^2\left(\sum_{k=-\infty}^{\infty}C_{n+k}\bar{C}_k\right)$$

由于

$$\sum_{n=-\infty}^{\infty}|B(n)|=\sigma^2\sum_{n=-\infty}^{\infty}\left|\sum_{k=-\infty}^{\infty}C_{n+k}C_k\right|$$

$$\leqslant\sigma^2\left(\sum_{k=\infty}^{\infty}|C_k|\right)^2<\infty$$

所以 $Y(n)$有连续谱密度 $f(\lambda)$, 且

$$f(\lambda)=\frac{1}{2\pi}\sum_{n=-\infty}^{\infty}B(n)e^{-in\lambda}$$

$$=\frac{\sigma^2}{2\pi}\sum_{n=-\infty}^{\infty}e^{-in\lambda}\left(\sum_{k=-\infty}^{\infty}C_{n+k}\bar{C}_k\right)$$

$$=\frac{\sigma^2}{2\pi}\left|\sum_{k=-\infty}^{\infty}C_ke^{-ik\lambda}\right|^2,\ -\pi\leqslant\lambda\leqslant\pi \tag{10}$$

事实上，当C_n满足$\sum_n |C_n|^2<\infty$时，(9)规定的Y_n也以(10)中的规定的$f(\lambda)$为谱密度（见例8)，只是此时(10)的级数是在$L^2[-\pi,\pi]$均方意义下收敛，而$f(\lambda)$也不一定是λ的连续函数.

［例3］ 若Y_T由§1例6规定，则其协方差函数为

$$B(t)=a^2e^{-2\mu t},\quad t>0$$

由于$\int_{-\infty}^{\infty}|B(t)|dt<\infty$，所以$Y_T$也有连续谱密度$f(\lambda)$，它可表为

$$\begin{aligned}f(\lambda)&=\frac{1}{2\pi}\int_{-\infty}^{\infty}e^{-i\lambda t}B(t)dt\\&=\frac{a^2}{2\pi}\left(\int_0^{\infty}e^{-(2\mu+i\lambda)t}dt+\int_{-\infty}^{0}e^{(2\mu-i\lambda)t}dt\right)\\&=\frac{2a^2\mu}{\pi(4\mu^2+\lambda^2)}\end{aligned}$$

［例4］ 若$X_T=\{X(n),n=0,\pm1,\cdots\}$为平稳正态马尔可夫时间序列，则由(1.23)，$\sum_n|B(n)|<\infty$，按(7)它也有连续谱密度$f(\lambda)$，且可表为

$$\begin{aligned}f(\lambda)&=\frac{1}{2\pi}\sum_{n=-\infty}^{\infty}B(n)e^{-in\lambda}\\&=\frac{1}{2\pi}\left(\sum_{n=0}^{\infty}a^nB(0)e^{-in\lambda}+\sum_{n=-\infty}^{-1}\overline{a^{-n}}B(0)e^{-in\lambda}\right)\\&=\frac{B(0)}{2\pi}\left(\frac{1}{1-ae^{-i\lambda}}+\frac{\bar{a}e^{i\lambda}}{1-\bar{a}e^{i\lambda}}\right)\\&=\frac{B(0)(1-|a|^2)}{2\pi\cdot(1-ae^{-i\lambda})(1-\bar{a}e^{i\lambda})}\end{aligned}\tag{11}$$

类似的，对连续参数平稳正态马尔可夫过程，若其协方差函数为

$$B(t)=e^{at}B(0),\ t>0,\ \mathrm{Re}\,a<0$$

则它有如下连续谱密度 $f(\lambda)$,

$$f(\lambda)=\frac{1}{2\pi}\int_{-\infty}^{\infty}B(t)e^{-i\lambda t}dt$$

$$=\frac{-\mathrm{Re}a}{\pi(a-i\lambda)(\bar{a}+i\lambda)}B(0) \tag{12}$$

[例 5] 若 X_T 为由 § 1 例 5 所规定的随机振动之和,即

$$X(t)=\sum_{n=-\infty}^{\infty}Z_n e^{i\omega_n t} \tag{13}$$

其中$\{Z_n\}$亦满足(1.10),则其协方差函数为

$$B(t)=\sum_{n=-\infty}^{\infty}\sigma_n^2 e^{i\omega_n t}$$

若取

$$F(\lambda)=\sum_{\omega_n<\lambda}\sigma_n^2 \tag{14}$$

则

$$B(t)=\int_{-\infty}^{\infty}e^{i\lambda t}dF(\lambda)$$

这个例子表明对于随机简谐振动的迭加而成的平稳过程(13),它的谱函数 $F(\lambda)$ 是由一个纯跳跃函数构成, $F(\lambda)$ 的跳跃点就是 $X(t)$ 各分量的振动(圆)频率 ω_n, 而跃度就是各分量振幅 Z_n 的方差 σ_n^2. 所以 $F(\lambda)$可看为各随机振动振幅方差随频率的分布函数,若不计常数因子之差,则可认为方差是与振幅平方同量纲的,即代表振动的能量或功率大小的,因此 $F(\lambda)$也是随机振动的功率随频率分布的分布函数,所以工程上也把 $F(\lambda)$称为**功率谱分布**,而其导数 $f(\lambda)$ 也称为**功率谱密度**.

[例 6] 若 $F(\lambda)$为任一左连续非降有界函数, $F(-\infty)=0$, ξ,η 为相互独立随机变量, ξ 以 $F(\lambda)/F(+\infty)$为其分布函数, η 为$[0,2\pi]$上均匀分布,令

$$Y(t)=\sqrt{F(+\infty)}\,e^{i(\xi t+\eta)}$$

则

$$E\{Y(t)\}=\sqrt{F(+\infty)}\int_{-\infty}^{\infty}\int_{0}^{2\pi}e^{i(tx+y)}\cdot d\left(\frac{F(x)}{F(+\infty)}\right)\frac{dy}{2\pi}=0$$

$$\begin{aligned}\Gamma(t+h,h)&=EY(t+h)\overline{Y(h)}\\&=F(+\infty)\int_{-\infty}^{\infty}\int_{0}^{2\pi}e^{itx}\cdot d\left(\frac{F(x)}{F(+\infty)}\right)\frac{dy}{2\pi}\\&=\int_{-\infty}^{\infty}e^{it\lambda}dF(\lambda)\end{aligned}$$

这就表明$\{Y(t)\}$是一个平稳过程. 这个例子告诉我们任一左连续非降有界函数都可作为某一平稳过程的谱函数, 而由此F按(3)规定的$B(t)$必是该平稳过程的协方差函数, 所以(3)不仅是$B(t)$作为协方差函数的必要条件, 它也是一个充分条件.

上面只是讨论了协方差函数的谱分解, 下面我们再来讨论平稳过程本身的谱分解. 为了讨论方便起见, 在下面两个定理中我们都假定平稳过程$\{X(t)\}$的均值为零. 因为当$X(t)$的均值m不为零时, 我们可以令$X_1(t)=X(t)-m$, 则$X_1(t)$就是均值为零的平稳过程了.

定理 3 设$X_T=\{X(n),\ n=0,\ \pm1,\cdots\}$为平稳随机序列, $E\{X(n)\}=0$, 则$X(n)$必可表为

$$X(n)=\int_{-\pi}^{\pi}e^{in\lambda}d\xi(\lambda) \tag{15}$$

其中$\{\xi(\lambda),-\pi\leqslant\lambda\leqslant\pi\}$若不计随机变量之差为由$X_T$唯一确定的满足下列条件的左连续正交增量过程.

(1) $E\{\xi(\lambda_2)-\xi(\lambda_1)\}=0;$ (16)

(2) 若$(\lambda_1,\lambda_1+\Delta\lambda_1)$与$(\lambda_2,\lambda_2+\Delta\lambda_2)$不相重迭, 则

$$E[\xi(\lambda_1+\Delta\lambda_1)-\xi(\lambda_1)]\overline{[\xi(\lambda_2+\Delta\lambda_2)-\xi(\lambda_2)]}=0 \tag{17}$$

(3) $E|\xi(\lambda_2)-\xi(\lambda_1)|^2=F(\lambda_2)-F(\lambda_1),\ \lambda_2>\lambda_1,$ (18)

其中$F(\lambda)$为X_T的谱函数.

定理 4 设 $X_T=\{X(t), -\infty<t<\infty\}$ 为均方连续平稳随机过程，$E\{X(t)\}=0$，则 $X(t)$ 必可表为

$$X(t)=\int_{-\infty}^{\infty} e^{it\lambda}d\zeta(\lambda) \tag{19}$$

其中 $\{\zeta(\lambda), -\infty<\lambda<\infty\}$ 若不计随机变量之差为由 X_T 唯一确定的满足下列条件的左连续正交增量过程，

(1) $E\{\zeta(\lambda_2)-\zeta(\lambda_1)\}=0$; (20)

(2) 若 $(\lambda_1, \lambda_1+\Delta\lambda_1)$ 与 $(\lambda_2, \lambda_2+\Delta\lambda_2)$ 不相重迭，则

$$E[\zeta(\lambda_1+\Delta\lambda_1)-\zeta(\lambda_1)]\overline{[\zeta(\lambda_2+\Delta\lambda_2)-\zeta(\lambda_2)]}=0 \tag{21}$$

(3) $E|\zeta(\lambda_2)-\zeta(\lambda_1)|^2=F(\lambda_2)-F(\lambda_1), \lambda_2\geqslant\lambda_1$, (22)

其中 $F(\lambda)$ 为 X_T 的谱函数.

由于上述两个定理的证明是相同的，所以我们把它放在一起叙述，而只对定理 4 进行证明.

[证明] 若 $F(\lambda)$ 为 X_T 的谱函数，记

$$H=\{\xi: E|\xi|^2<\infty, \xi \text{ 为随机变量}\} \tag{23}$$

$$L_2(dF)=\{\varphi: \varphi(\lambda) \text{ 为}(-\infty, \infty)\text{上复值函数}, \int_{-\infty}^{\infty}|\varphi(\lambda)|^2dF(\lambda)<\infty\} \tag{24}$$

如在 H 中以 $E\{\xi\bar{\eta}\}$ 表 ξ, η 的内积 (ξ, η)，并把以概率 1 相等的随机变量看为是同一个随机变量，则 H 是一个希尔伯特空间，同样 $L_2(dF)$ 也是一个希尔伯特空间，它的内积是

$$(\varphi, \psi)=\int_{-\infty}^{\infty}\varphi(\lambda)\overline{\psi(\lambda)}dF(\lambda)$$

我们是靠在 $L_2(dF)$ 与 H 建立某种对应来证明这个定理的.

(i) 首先建立 $L_2(dF)$ 到 H 中的如下对应 f：对

$$\varphi(\lambda)=\sum_{j=1}^{n}c_je^{it_j\lambda}\in L_2(dF)$$

$$f(\varphi)=\sum_{j=1}^{n}c_jX(t_j) \tag{25}$$

则

$$E\{f(\varphi)\}=E\left\{\sum_{j=1}^{n}c_jX(t_j)\right\}=0$$

且当 $\psi(\lambda)=\sum_{k=1}^{m}d_ke^{it'_k\lambda}$ 时，利用 X_T 协方差函数 $B(t)$ 的谱分解有：

$$(f(\varphi),f(\psi))=E\left\{\left(\sum_{j=1}^{n}c_jX(t_j)\right)\cdot\overline{\left(\sum_{k=1}^{m}d_kX(t'_k)\right)}\right\}$$

$$=\sum_{j,k=1}^{n,m}c_j\bar{d}_kB(t_j-t'_k)$$

$$=\sum_{j,k=1}^{n,m}c_j\bar{d}_k\int_{-\infty}^{\infty}e^{i(t_j-t'_k)\lambda}dF(\lambda)$$

$$=\int_{-\infty}^{\infty}\left(\sum_{j=1}^{n}c_je^{it_j\lambda}\right)\overline{\left(\sum_{k=1}^{m}d_ke^{it'_k\lambda}\right)}dF(\lambda)$$

$$=\int_{-\infty}^{\infty}\varphi(\lambda)\overline{\psi(\lambda)}dF(\lambda)$$

$$=(\varphi,\psi)$$

利用这一等式，首先可以说明 f 是一意的，即当 $\varphi(\lambda)=\sum_{j=1}^{n}c_je^{it_j\lambda}$ $=\sum_{k=1}^{m}d_ke^{it'_k\lambda}$ 时，必有 $f\left(\sum_{j=1}^{n}c_je^{it_j\lambda}\right)=\sum_{j=1}^{n}c_jX(t_j)=\sum_{k=1}^{m}d_kX(t'_k)$ $=f\left(\sum_{k=1}^{m}d_ke^{it'_k\lambda}\right)$. 其次 f 是线性的，并且 f 还是保持内积不变的，而当 $L_2(dF)$ 中的一列元素 φ_n 满足下列条件时，

$$\lim_{n\to\infty}\|\varphi_n-\varphi\|=0$$

对H中的元素列$\{f(\varphi_n)\}$就有

$$E|f(\varphi_n)-f(\varphi_m)|^2=\|f(\varphi_n)-f(\varphi_m)\|^2$$
$$=\|\varphi_n-\varphi_m\|^2\to 0$$

由H的完备性,必有H中元素ξ,使

$$\lim_{n\to\infty}E|\xi-f(\varphi_n)|^2=0$$

我们就可令

$$\xi=f(\varphi)$$

不难说明,这样规定$f(\varphi)$是完全确定的,即$f(\varphi)$不随收敛于φ的$\{\varphi_n\}$不同而改变,而且当$\eta=f(\psi)$时,仍然成立

$$E\eta=0$$

$$E\{\xi\bar{\eta}\}=\int_{-\infty}^{\infty}\varphi(\lambda)\overline{\psi(\lambda)}dF$$

这样我们就在$L_2(dF)$中形为$\varphi(\lambda)=\sum_{j=1}^{n}c_je^{it_j\lambda}$及其均方极限的所有元素上建立了到$H$中的映照$f$,且$f$保持内积不变

$$E\{f(\varphi)\}=0$$
$$(f(\varphi),f(\psi))=(\varphi,\psi)$$

由于$\{e^{it\lambda},-\infty<t<\infty\}$在$L_2(dF)$中的完备性,即对$L_2(dF)$中任一函数$\varphi$都可选一列$\varphi_k=\sum_{j=1}^{n_k}c_j^{(k)}e^{it_j^{(k)}\lambda}$,使

$$\lim_{k\to\infty}\int_{-\pi}^{\pi}|\varphi(\lambda)-\varphi_k(\lambda)|^2dF(\lambda)=0$$

所以f是$L_2(dF)$到H中的映照,且保持内积不变.

(ii) 对任何实数a,b,以$\chi_{[a,b)}(\lambda)$表区间$[a,b)$的示性函数,则

$$\chi_{[a,b)}(\lambda)\in L_2(dF)$$

$$\chi_{(-\infty,b)}(\lambda)\in L_2(dF)$$

令

$$\zeta(b)=f(\chi_{(-\infty,b)})\tag{26}$$

则

$$E\{\zeta(\lambda)\}=0$$

且当$(\lambda_1,\lambda_1+\Delta\lambda_1)$与$(\lambda_2,\lambda_2+\Delta\lambda_2)$不相交时，

$$\begin{aligned}&E(\zeta(\lambda_1+\Delta\lambda_1)-\zeta(\lambda_1))\overline{(\zeta(\lambda_2+\Delta\lambda_2)-\zeta(\lambda_2))}\\&=E\{f(\chi_{[\lambda_1,\lambda_1+\Delta\lambda_1)}(\lambda))\overline{f(\chi_{[\lambda_2,\lambda_2+\Delta\lambda_2)}(\lambda)}\}\\&=\int_{-\infty}^{\infty}\chi_{[\lambda_1,\lambda_1+\Delta\lambda_1)}(\lambda)\chi_{[\lambda_2,\lambda_2+\Delta\lambda_2)}(\lambda)dF(\lambda)\\&=0\end{aligned}$$

即$\{\zeta(\lambda),-\infty<\lambda<\infty\}$为正交增量过程，且当$\lambda_2>\lambda_1$时

$$\begin{aligned}E|\zeta(\lambda_2)-\zeta(\lambda_1)|^2&=E|f(\chi_{[\lambda_1,\lambda_2)}(\lambda))|^2\\&=\int_{-\infty}^{\infty}|\chi_{[\lambda_1,\lambda_2)}(\lambda)|^2dF(\lambda)\\&=F(\lambda_2)-F(\lambda_1)\end{aligned}$$

所以$\{\zeta(\lambda)\}$满足(20)——(22)且为左连续的。

(iii) 由于$\{\zeta(\lambda)\}$满足(20)——(22)且为左连续的，所以我们可以对$\varphi\in L_2(dF)$规定如下随机积分

$$\int_{-\infty}^{\infty}\varphi(\lambda)d\zeta(\lambda)$$

并且满足

$$E\left\{\int_{-\infty}^{\infty}\varphi(\lambda)d\zeta(\lambda)\right\}=0$$

$$E\left\{\int_{-\infty}^{\infty}\varphi(\lambda)d\zeta(\lambda)\overline{\int_{-\infty}^{\infty}\psi(\lambda)d\zeta(\lambda)}\right\}=\int_{-\infty}^{\infty}\varphi(\lambda)\overline{\psi(\lambda)}dF(\lambda)$$

特别当$\varphi(\lambda)$为λ的连续函数时

$$\int_{-\infty}^{\infty} \varphi(\lambda) d\zeta(\lambda) = \lim_{N\to\infty} \int_{-N}^{N} \varphi(\lambda) d\zeta(\lambda)$$

$$= \lim_{N\to\infty} \lim_{\max_j |\lambda_{j+1}^{(N)} - \lambda_j^{(N)}| \to 0} \sum_{j=1}^{K_N - 1} \varphi(\tilde{\lambda}_j^{(N)}) [\zeta(\lambda_{j+1}^{(N)}) - \zeta(\lambda_j^{(N)})] \quad (27)$$

其中

$$\lambda_j^{(N)} \leqslant \tilde{\lambda}_j^{(N)} \leqslant \lambda_{j+1}^{(N)}$$

$$-N = \lambda_1^{(N)} \leqslant \lambda_2^{(N)} \leqslant \cdots \leqslant \lambda_{k_N}^{(N)} = N$$

而(27)中的极限都是H中的均方收敛极限，利用(27)不难得出

$$E\left\{\left(\int_{-\infty}^{\infty} \varphi(\lambda) dz(\lambda)\right) \overline{X(t)}\right\}$$

$$= \lim_{N\to\infty} \lim_{\max_j |\lambda_{j+1}^{(N)} - \lambda_j^{(N)}| \to 0} \sum_{j=1}^{k_N - 1} \varphi(\tilde{\lambda}_j^N) E\{[\zeta(\lambda_{j+1}^{(N)}) - \zeta(\lambda_j^{(N)})] \overline{X(t)}\}$$

$$= \lim_{N\to\infty} \lim_{\max_j |\lambda_{j+1}^{(N)} - \lambda_j^{(N)}| \to 0} \sum_{j=1}^{k_N - 1} \varphi(\tilde{\lambda}_j^{(N)}) \int_{-\infty}^{\infty} \chi_{[\lambda_j^{(N)}, \lambda_{j+1}^{(N)})}(\lambda) e^{-it\lambda} dF(\lambda)$$

$$= \int_{-\infty}^{\infty} \varphi(\lambda) e^{-it\lambda} dF \quad (28)$$

(iv) 取 $\varphi(\lambda) = e^{it\lambda}$ 利用(28)可得：

$$E\left| X(t) - \int_{-\infty}^{\infty} e^{it\lambda} d\zeta(\lambda) \right|^2$$

$$= E|X(t)|^2 - 2\mathrm{Re} E\left\{ X(t) \cdot \int_{-\infty}^{\infty} e^{it\lambda} d\zeta(\lambda) \right\}$$

$$+ E\left| \int_{-\infty}^{\infty} e^{it\lambda} d\zeta(\lambda) \right|^2$$

$$= B(0) - 2\mathrm{Re} \int_{-\infty}^{\infty} dF(\lambda) + \int_{-\infty}^{\infty} dF(\lambda)$$

$$= 0$$

即

$$X(t) = \int_{-\infty}^{\infty} e^{it\lambda} d\zeta(\lambda)$$

(v) 若 $X(t)$ 有表示式(19)，且 $\zeta(\lambda)$ 满足(20)——(22)，则因 $\{e^{it\lambda}, -\infty<t<\infty\}$ 在 $L_2(dF)$ 中完备，对任何实数 a，b 必有 $\varphi_n(\lambda)=\sum_{k=1}^{k_n} c_k e^{it_k\lambda}$ 满足

$$\lim_{n\to\infty}\int_{-\infty}^{\infty}|\varphi_n(\lambda)-\chi_{(a,b)}(\lambda)|^2 dF(\lambda)=0$$

这时利用 $\zeta(\lambda)$ 为正交增量过程，必有

$$\begin{aligned}
&\lim_{n\to\infty}\sum_{k=1}^{k_n} c_k X(t_k)\\
&=\lim_{n\to\infty}\sum_{k=1}^{k_n} c_k\int_{-\infty}^{\infty} e^{it_k\lambda}d\zeta(\lambda)\\
&=\lim_{n\to\infty}\int_{-\infty}^{\infty}\varphi_n(\lambda)d\zeta(\lambda)\\
&=\int_{-\infty}^{\infty}\chi_{(a,b)}(\lambda)d\zeta(\lambda)\\
&=\zeta(b)-\zeta(a)
\end{aligned}\tag{29}$$

所以 $\zeta(\lambda)$ 的增量由 X_T 唯一确定，且不计随机变量之差，$\zeta(\lambda)$ 本身也是唯一确定的. 定理证毕.

对于实的平稳过程更有如下定理.

定理 5 (i) 若 $X_T=\{X(n), -\infty<n<\infty\}$ 是均值为零的实的平稳随机序列，则 $X(n)$ 有如下表示

$$X(n)=\int_0^{\pi}\cos n\lambda dU(\lambda)+\int_0^{\pi}\sin n\lambda dV(\lambda)\tag{30}$$

其中 $\{U(\lambda), -\pi<\lambda<\pi\}$，$\{V(\lambda), -\pi<\lambda<\pi\}$，若不计随机变量之差，为由 X_T 唯一确定的满足下列条件的左连续正交增量过程

(1) $E\{U(\lambda)\}=E\{V(\lambda)\}=0$; (31)

(2) 若 $\lambda_1\leqslant\lambda_3\leqslant\lambda_2\leqslant\lambda_4$，则

$$\begin{aligned}
&E\{(U(\lambda_2)-U(\lambda_1))(U(\lambda_4)-U(\lambda_3))\}\\
&=E\{(V(\lambda_2)-V(\lambda_1))(V(\lambda_4)-V(\lambda_3))\}
\end{aligned}$$

$$=\int_{\lambda_1}^{\lambda_2} dF_1(\lambda) \tag{32}$$

$$E(U(\lambda_2)-U(\lambda_1))(V(\lambda_4)-V(\lambda_3))=0 \tag{33}$$

其中 $F_1(\lambda)$ 为 $X(n)$ 协方差函数 $B(n)$ 如下谱分解中的谱函数,

$$B(n)=\int_0^{\pi}\cos n\lambda dF_1(\lambda)$$

(ii) 若 $X_T=\{X(t), -\infty<t<\infty\}$ 是均值为零的实的均方连续平稳过程, 则 $X(t)$ 有如下表示

$$X(t)=\int_0^{\infty}\cos t\lambda dU(\lambda)+\int_0^{\infty}\sin t\lambda dV(\lambda) \tag{34}$$

其中 $\{U(\lambda), -\infty<\lambda<\infty\}$, $\{V(\lambda), -\infty<\lambda<\infty\}$ 若不计随机变量之差为由 X_T 唯一决定的满足(31)——(33)的左连续正交增量过程.

[证明] 我们只证明(ii), 对于(i)证明是完全同样的. 若 X_T 的协方差函数 $B(t)$ 有谱分解

$$B(t)=\int_0^{\infty}\cos t\lambda dF_1(\lambda)$$

$$=\int_{-\infty}^{\infty}e^{it\lambda}dF(\lambda)$$

则由定理 2 可知当 $\lambda_1, \lambda_2>0$ 时,

$$F_1(\lambda_2)-F_1(\lambda_1)=2[F(\lambda_2)-F(\lambda_1)]$$

$$=2[F(-\lambda_1+0)-F(-\lambda_2+0)]$$

按定理 4

$$X(t)=\int_{-\infty}^{\infty}e^{i\lambda t}d\xi(\lambda)$$

且

$$E|\xi(\lambda_2)-\xi(\lambda_1)|^2=F(\lambda_2)-F(\lambda_1)$$

由于 $X(t)$ 是实的, 所以

$$X(t)=\overline{X(t)}=\int_{-\infty}^{\infty}e^{-it\lambda}d\overline{\xi(\lambda)}$$

因此由(29)中$\xi(\lambda)$由$X(t)$唯一确定可知

$$\xi(\lambda_2)-\xi(\lambda_1)=\overline{\xi(-\lambda_1+0)}-\overline{\xi(-\lambda_2+0)} \tag{35}$$

并且当$\lambda_1,\lambda_2,\lambda_3,\lambda_4\geqslant 0$时,由(21)有

$$\begin{aligned}&E\{(\xi(\lambda_2)-\xi(\lambda_1))(\xi(\lambda_4)-\xi(\lambda_3))\}\\=&E[\xi(\lambda_2)-\xi(\lambda_1)]\overline{[\xi(-\lambda_3+0)-\xi(-\lambda_4+0)]}\\=&0\end{aligned} \tag{36}$$

若$\xi_1(\lambda)$,$\xi_2(\lambda)$分别为$\xi(\lambda)$的实部和虚部,将复值$\xi(\lambda)$表为

$$\xi(\lambda)=\xi_1(\lambda)+i\xi_2(\lambda)$$

则由(22)及取(36)的实部和虚部分别有

$$\begin{aligned}&E[\xi_1(\lambda_2)-\xi_1(\lambda_1)]^2=E[\xi_2(\lambda_2)-\xi_2(\lambda_1)]^2\\=&\frac{1}{2}E|\xi(\lambda_2)-\xi(\lambda_1)|^2=\frac{1}{2}[F(\lambda_2)-F(\lambda_1)]\end{aligned} \tag{37}$$

$$\begin{aligned}&E(\xi_1(\lambda_2)-\xi_1(\lambda_1))(\xi_2(\lambda_4)-\xi_2(\lambda_3))\\=&-E[\xi_2(\lambda_2)-\xi_2(\lambda_1)][\xi_1(\lambda_4)-\xi_1(\lambda_3)]\end{aligned}$$

又由

$$\begin{aligned}&-E[\xi_1(\lambda_2)-\xi_1(\lambda_1)][\xi_2(\lambda_4)-\xi_2(\lambda_3)]\\&+E[\xi_2(\lambda_2)-\xi_2(\lambda_1)][\xi_1(\lambda_4)-\xi_1(\lambda_3)]\\&=\mathrm{Im}E[\xi(\lambda_2)-\xi(\lambda_1)]\overline{[\xi(\lambda_4)-\xi(\lambda_3)]}=0\end{aligned}$$

可得

$$E(\xi_1(\lambda_2)-\xi_1(\lambda_1))(\xi_2(\lambda_4)-\xi_2(\lambda_3))=0 \tag{38}$$

因为X_T是实的,

$$\begin{aligned}X(t)&=\mathrm{Re}\int_{-\infty}^{\infty}e^{it\lambda}d\xi(\lambda)\\&=\int_{-\infty}^{\infty}\cos\lambda t d\xi_1(\lambda)-\int_{-\infty}^{\infty}\sin\lambda t d\xi_2(\lambda)\\&=\int_0^{\infty}\cos\lambda t dU(\lambda)+\int_0^{\infty}\sin\lambda t dV(\lambda)\end{aligned}$$

其中U,V由于(35)可取为

$$U(\lambda)=2\xi_1(\lambda),V(\lambda)=-2\xi_2(\lambda)$$

再从(37)，(38)就不难推出 U, V 满足(31)——(33)，至于 U, V 由 X_T 唯一确定，其证明也是类似的.

由定理证明过程中的(29)，我们可以看到 $\zeta(b)-\zeta(a)$ 是由 $X(t)$ 的线性组合的均方极限来规定的，当 $X(t)$ 为实过程时，其随机谱函数 $U(\lambda)$, $V(\lambda)$ 也是实的，而逼近 $U(\lambda_2)-U(\lambda_1)$, $V(\lambda_2)-V(\lambda_1)$ 的线性组合亦可取为实系数的，这时当 $X(t)$ 为实正态过程时，$U(\lambda)$, $V(\lambda)$ 也可取为实正态过程. 若我们能适当地规定复正态过程，只要它保线性变换及均方收敛下分布正态性不变，那么对复平稳正态过程的随机谱函数也可取为复的正态过程，而 $\zeta(\lambda)$ 增量的不相关性就变为独立性，这样就有下列推论.

推论 $X_T=\{X(t),\ -\infty<t<\infty\}$ 为均方连续平稳正态过程的充要条件是其谱分解式(19)中的 $\zeta(\lambda)$ 可取为正态独立增量过程.

上面三个定理都是把平稳过程表示为对正交增量过程的随机积分，(15)，(19)，(30)，(34)都称为随机过程 X_T 的**谱分解**，其中的 $\zeta(\lambda)$ 和 $U(\lambda)$, $V(\lambda)$ 也称为过程 X_T 的**随机谱函数**.

谱分解式(15)，(19)，(30)，(34)就是§1例4，5的推广，因为对于随机积分

$$X(t)=\int e^{it\lambda}d\zeta(\lambda)$$

可以用如下的和式来逼近

$$Y(t)=\sum_{j=1}^{n}e^{it\lambda_j}[\zeta(\lambda_{j+1})-\zeta(\lambda_j)]$$

而对上述 $Y(t)$，它就是频率为 λ_j 的具有随机振幅 $\zeta(\lambda_{j+1})-\zeta(\lambda_j)$ 的随机简谐振动的迭加，而由于 $\zeta(\lambda)$ 是正交增量过程，所以 $Y(t)$ 中各随机振幅 $\zeta(\lambda_{j+1})-\zeta(\lambda_j)$ 是互不相关的，而对 λ 的分割加密，$Y(t)$ 均方收敛于 $X(t)$ 这一点就表明对一般的平稳过程，不能只

把它看为仅是有限个频率随机振动的迭加，而应该是无限个各种不同频率随机振动的迭加，随机积分就是这种迭加的数学描述。

对离散参数的平稳序列，频率变化范围是$[-\pi,\pi]$，对连续参数过程频率可在$(-\infty,\infty)$上变化，而对实的过程只需考虑频率为正的情况，过程$\xi(\lambda)$在区间$[\lambda_1,\lambda_2)$上的增量$\xi(\lambda_2)-\xi(\lambda_1)$，就是频率在$[\lambda_1,\lambda_2)$上各随机振动振幅的迭加，所以

$$E|\xi(\lambda_2)-\xi(\lambda_1)|^2=F(\lambda_2)-F(\lambda_1)$$

就是频率在$[\lambda_1,\lambda_2)$上诸振动的功率，而

$$E|X(t)|^2=B(0)=\int dF(\lambda)$$

正表明反映变化过程$X(t)$的总功率$E|X(t)|^2$是其各频率上功率$dF(\lambda)$的总和，这就是谱分解定理的物理解释．正由于平稳过程的谱函数有这种物理含义，所以从谱分解出发来处理平稳过程各种问题即频率域处理方法就成为一种重要的方法了。

此外，谱分解式

$$X(t)=\int e^{it\lambda}d\xi(\lambda)$$

是用随机积分来规定的，正如随机积分不能看作样本函数的斯梯杰积分一样，它也不能被理解为$X(t)$的样本函数有同样的富利埃展开，(15)，(19)，(30)，(34)只能表明$X(t)$是无数简谐振动的“均方”迭加，而对$X(t)$本身的样本函数可以是不具备这种分解的，因为对$X(t)$本身来说，它的样本函数如下积分$\int_{-\infty}^{\infty}|X(t)|^2dt$往往是无限的，所以是难于建立它的富利埃展开式的。

对于平稳过程来说，谱分解定理作为一个基本定理，它的重要性不仅在于它的物理含义，而且它也为我们提供了分析平稳过程的一个数学工具．因为在谱分解定理的证明中，我们在$L_2(dF)$到H建立了如下的映照f

$$L_2(dF)\ni e^{it\lambda}\xrightarrow{f}X(t)\in H$$

对一般的 $\varphi\in L_2(dF)$，不难看出

$$f(\varphi)=\int\varphi(\lambda)d\xi(\lambda) \tag{39}$$

而在 f 之下 $L_2(dF)$ 的象全体就是 H 中 $\{X(t),t\in T\}$ 线性组合及其均方极限的全体，若以 $H_T(X)$ 表示这个集合，那么 f 就是 $L_2(dF)$ 到 $H_T(X)$ 上的映照，且保持内积不变，即

$$(f(\varphi),f(\psi))_H=(\varphi,\psi)_{L_2(dF)} \tag{40}$$

所以这个映照也必是一一的(不计随机变量以概率 1 相等的差别)．这个映照的逆映照 f^{-1} 就把 $H_T(X)$ 照到 $L_2(dF)$，由于保持内积不变，所以 $H_T(X)$ 上关于 X_T 的许多分析问题，可以在函数空间 $L_2(dF)$ 上来进行，这样利用泛函分析或富利埃分析中的一些方法和结论，就可以从 $L_2(dF)$ 中来帮助解决 $H_T(X)$ 中的概率统计问题，当然从概率统计的直观出发解决 $H_T(X)$ 中某些概率统计问题，也会同时获得 $L_2(dF)$ 中相应的结果，所以谱分解及随之而确立的 $L_2(dF)$ 与 $H_T(X)$ 间的映照成为研究平稳过程的一种重要方法．

［例 7］ 若 $\{X(n),n=0,\pm1,\cdots\}$ 为例 1 中的零均值互不相关序列，又实数列 $\{c_k\}$ 满足

$$\sum_{k=-\infty}^{\infty}|c_k|^2<\infty$$

按§ 1例 2，可规定平稳过程

$$Y(n)=\sum_{k=-\infty}^{\infty}c_kX(n-k) \tag{41}$$

因为 $\{X(n)\}$ 是平稳过程，若它的谱分解为

$$X(n)=\int_{-\pi}^{\pi}e^{in\lambda}d\xi_X(\lambda)$$

则由例 1, $\xi_X(\lambda)$满足

$$E|\xi_X(\lambda_2)-\xi_X(\lambda_1)|^2=\frac{\lambda_2-\lambda_1}{2\pi}\sigma^2$$

这时利用 $L_2(dF)$与 $H_T(X)$的一一对应，又由(41)在 $H_T(X)$中均方收敛，所以

$$X(n-k)\longleftrightarrow e^{i(n-k)\lambda}$$

$$c_kX(n-k)\longleftrightarrow c_ke^{i(n-k)\lambda}$$

$$Y(n)=\sum_{k=-\infty}^{\infty}c_kX(n-k)\longleftrightarrow\sum_{k=-\infty}^{\infty}c_ke^{i(n-k)\lambda}$$

$$=e^{in\lambda}(\sum_{k=-\infty}^{\infty}c_ke^{-ik\lambda})=\cdots$$

这样，由(39)可得

$$Y(n)=f(e^{in\lambda}(\sum_{k=-\infty}^{\infty}c_ke^{-ik\lambda}))$$

$$=\int_{-\pi}^{\pi}e^{in\lambda}(\sum_{k=-\infty}^{\infty}c_ke^{-ik\lambda})d\xi_X(\lambda)$$

$$=\int_{-\pi}^{\pi}e^{in\lambda}d\xi_Y(\lambda)$$

因此 $Y(n)$的随机谱函数可取为

$$\xi_Y(\lambda)=\int_{-\pi}^{\lambda}(\sum_{u=-\infty}^{\infty}c_ke^{-ik\mu})d\xi_X(\mu)$$

而 $Y(n)$的谱函数也不难由此推出为

$$F(\lambda)=\frac{\sigma^2}{2\pi}\int_{-\pi}^{\lambda}\left|\sum_{u=-\infty}^{\infty}c_ke^{-ik\mu}\right|^2d\mu$$

即 $Y(n)$是存在如下谱密度

$$f(\lambda)=\frac{\sigma^2}{2\pi}\left|\sum_{u=-\infty}^{\infty}c_ke^{-ik\lambda}\right|^2$$

由 § 1 知道，平稳过程 $X_T=\{X(t),\ -\infty<t<\infty\}p$ 次均方可

微的充要条件是相关函数 $B(\tau)$ 在 $\tau=0$ 时 $2p$ 次可微，类似于特征函数可微性与分布函数存在矩之间的关系，又可推出 X_T 为 p 次均方可微的充要条件是

$$\int_{-\infty}^{\infty}\lambda^{2p}dF(\lambda)<\infty \tag{42}$$

若 X_T 为 p 次均方可微，则当 $q\leqslant p$ 时，由(1.16)及(42)式成立

$$\begin{aligned}E\{X^{(q)}(t)X^{(q)}(s)\}&=(-1)^q\frac{\partial^{2q}}{\partial t^q\partial s^q}B(t-s)\\&=\int_{-\infty}^{\infty}\lambda^{2q}e^{i(t-s)\lambda}dF(\lambda)\end{aligned} \tag{43}$$

所以$\{X^{(q)}(t),-\infty<t<\infty\}$也是一个均方连续平稳过程，若它的谱函数为 $F_q(\lambda)$，则由(43)

$$F_q(\lambda_2)-F_q(\lambda_1)=\int_{\lambda_1}^{\lambda_2}\lambda^{2q}dF(\lambda) \tag{44}$$

若 $X(t)$，$X^{(q)}(t)$有如下谱分解

$$X(t)=\int_{-\infty}^{\infty}e^{it\lambda}d\zeta(\lambda)$$

$$X^{(q)}(t)=\int_{-\infty}^{\infty}e^{it\lambda}d\zeta_q(\lambda)$$

则不难说明

$$\zeta_q(\lambda_2)-\zeta_q(\lambda_1)=\int_{\lambda_1}^{\lambda_2}(i\lambda)^q d\zeta(\lambda)$$

§3. 线性系统中的平稳过程

一、线性时不变系统

在工业生产和科学实验中遇到的各类随机过程问题中，往往是与“系统”相联系的，所谓系统就是能对各种“输入”按一定的规则产生“输出”，而研究系统的各种特性就是讨论输入、输出间的联系及其各种指标间的相互关系．例如各种自动控制装置都是系

统，进入这些装置的信号都是输入，而由这些装置随之而产生的控制信号或动作就是输出，在研究大气湍流对飞机的影响时，飞机本身就是一个系统，湍流的作用是输入，飞机各部分的振动就是输出．在电路分析中，电路本身是一个系统，电路中的电子热运动——热噪声就是输入，而电路输出端的电压或电流波动就是输出．

为了研究输入、输出间的数量关系，必须从数量上来描述输入 $x(t)$ 与输出 $y(t)$ 间的依赖关系．常用的可以是用微分或积分关系式；例如

$$\sum_{j=0}^{n} a_j y^{(n-j)}(t) = \sum_{k=0}^{m} b_k x^{(m-k)}(t) \tag{1}$$

或者

$$y(t)=\int_{-\infty}^{\infty} h(t,s)x(s)ds \tag{2}$$

在实际遇到的各类系统中，较为简单而又重要的一类是线性时不变系统，所谓线性时不变系统就是指它满足迭加原理，并且输出对输入的依赖关系不随时间推移而改变的，即当对应 $x_1(t)$，$x_2(t)$ 的输出分别为 $y_1(t)$，$y_2(t)$ 时，则对任意常数 α，β 对应 $\alpha x_1(t)+\beta x_2(t)$ 的输出为 $\alpha y_1(t)+\beta y_2(t)$，对应 $x_1(t+\tau)$ 的输出为 $y_1(t+\tau)$．若输出、输入用如下关系式表示

$$y(t)=L[x(t)]$$

则线性时不变系统就是满足下列条件的系统

$$L[\alpha x_1(t)+\beta x_2(t)]=\alpha L[x_1(t)]+\beta L[x_2(t)] \tag{3}$$

$$L[x(t+\tau)]=y(t+\tau) \tag{4}$$

对线性时不变系统有下列引理。

引理 1　对线性时不变系统 L，若其输入 $x(t)=e^{i\omega t}$，则其输

出是

$$y(t)=\Phi(\omega)e^{i\omega t} \tag{5}$$

[证明] 若

$$L[e^{i\omega t}]=y(t)$$

由(4)

$$L[e^{i\omega(t+\tau)}]=y(t+\tau)$$

因为 $e^{i(t+\tau)\omega}=e^{i\tau\omega}\cdot e^{it\omega}$，由(3)

$$L[e^{i\tau\omega}\cdot e^{it\omega}]=e^{i\tau\omega}y(t)$$

所以

$$y(t+\tau)=e^{i\tau\omega}y(t) \tag{6}$$

记

$$y(0)=L[e^{i\omega t}]|_{t=0}=\Phi(\omega)$$

则在(6)中令 $t=0$ 有

$$y(\tau)=\Phi(\omega)e^{i\tau\omega}$$

引理证毕.

上述引理表明，对线性时不变系统，若输入是 $e^{i\omega t}$，则输出仍为同一频率的函数，但振幅和相位有一修正，所以(5)中的 $\Phi(\omega)$ 称为系统的**频率响应**或**频率特性**，它一般是一个复数，可表为

$$\Phi(\omega)=A(\omega)e^{i\theta(\omega)}$$

其中 $A(\omega)=|\Phi(\omega)|$ 称为系统的**振幅特性**，$\theta(\omega)$ 称为系统的**相位特性**，所以对复正弦波 $e^{it\omega}$ 输入，它的输出的振幅衰减了一个因子 $A(\omega)$，相位相差了 $\theta(\omega)$.

当输入 $x(t)$ 为一个平方可积函数时

$$\int_{-\infty}^{\infty}|x(t)|^2dt<\infty$$

则从富利埃积分理论可知，$x(t)$ 可表为富利埃积分

$$x(t)=\frac{1}{2\pi}\int_{-\infty}^{\infty}X(\omega)e^{it\omega}d\omega \tag{7}$$

其中

$$X(\omega)=\int_{-\infty}^{\infty}x(t)e^{-it\omega}dt \tag{8}$$

又称为 $x(t)$ 的频谱. 对于(7), 我们也可认为 $x(t)$ 是下列形式 $x_n(t)$ 的极限

$$x_n(t)=\frac{1}{2\pi}\sum_j X(\omega_j)\Delta\omega_j\cdot e^{it\omega_j}$$

而 $x_n(t)$ 是 $e^{it\omega_j}$ 的线性组合. 由(3), (5)

$$y_n(t)=L[x_n(t)]=\frac{1}{2\pi}\sum_j X(\omega_j)\Delta\omega_j L[e^{it\omega_j}]$$

$$=\frac{1}{2\pi}\sum_j X(\omega_j)\Delta\omega_j\cdot\Phi(\omega_j)e^{it\omega_j}$$

若 L 还保持连续性, 当 $x_n(t)$ 收敛于 $x(t)$ 时 $y_n(t)$ 的极限也收敛于 $x(t)$ 的输出 $y(t)$, 则

$$y(t)=L[x(t)]=L[\lim_{n\to\infty}x_n(t)]$$

$$=\lim_{n\to\infty}L[x_n(t)]$$

$$=\lim_{n\to\infty}\{\frac{1}{2\pi}\sum_j X(\omega_j)\Delta\omega_j\Phi(\omega_j)e^{it\omega_j}\}$$

$$=\frac{1}{2\pi}\int_{-\infty}^{\infty}\Phi(\omega)X(\omega)e^{i\omega t}d\omega \tag{9}$$

当然对 $y(t)$ 亦有类似于(7)的富利埃积分表示式

$$y(t)=\frac{1}{2\pi}\int_{-\infty}^{\infty}Y(\omega)e^{i\omega t}d\omega \tag{10}$$

比较(9), (10)就可得 $y(t)$ 的频谱 $Y(\omega)$ 为

$$Y(\omega)=\Phi(\omega)X(\omega) \tag{11}$$

因为 $X(\omega)$ 和 $Y(\omega)$ 分别完全表征了 $x(t)$ 和 $y(t)$, 所以由 $\Phi(\omega)$ 就完全确定了系统的输入输出间的关系, 即线性时不变系统可以由

它的频率特征完全确定．当我们从频率特征出发来研究系统的种种特性时又称为系统的**频域分析**．

当然从(11)出发也可直接求 $x(t)$，$y(t)$的依赖关系，若 $\Phi(\omega)$ 本身是平方可积的

$$\int |\Phi(\omega)|^2 d\omega < \infty$$

而 $h(t)$满足下列关系式，

$$h(t)=\frac{1}{2\pi}\int_{-\infty}^{\infty}\Phi(\omega)e^{i\omega t}d\omega \tag{12}$$

$$\Phi(\omega)=\int_{-\infty}^{\infty}h(t)e^{-it\omega}dt \tag{13}$$

则由(11)，利用富利埃变换对函数卷积的公式就有

$$\begin{aligned} y(t)&=\frac{1}{2\pi}\int_{-\infty}^{\infty}Y(\omega)e^{it\omega}d\omega \\ &=\frac{1}{2\pi}\int_{-\infty}^{\infty}\Phi(\omega)X(\omega)e^{it\omega}d\omega \\ &=\int_{-\infty}^{\infty}h(t-\tau)x(\tau)d\tau \end{aligned} \tag{14}$$

公式(14)就是从时间域上联系输入 $x(t)$和输出 $y(t)$的关系式，当我们形式地把表示脉冲的 $\delta(\tau)$作为输入代入(14)时，就有

$$\begin{aligned} y(t)&=\int_{-\infty}^{\infty}h(t-\tau)\delta(\tau)d\tau \\ &=h(t) \end{aligned}$$

所以 $h(t)$就是输入尖脉冲时系统的输出，它也称为系统的**脉冲响应**．由(12)，(13)可见，系统的频率响应和脉冲响应是互为富利埃变换且相互单值确定的，(11)，(14)表明它们都能完全确定系统输入输出间的依赖关系．在不同的问题中，可依据问题的条件和不同要求，分别采用(11)或(14)来解决问题．

例如，若对输入 $x(t)$进行微分的装置 L，则输出

$$y(t)=L[x(t)]=x'(t)$$

不难验证，它也满足(3)，(4)是线性时不变的，由于

$$\Phi(\omega)=L[e^{i\omega t}]|_{t=0}=i\omega$$

所以它的频率响应为 $i\omega$. 这时由于 $\Phi(\omega)=i\omega$ 不是平方可积的，在普通函数中找不到使(14)成立的 $h(t)$，所以在这个例子中频域分析有它的方便之处. 类似的，当系统 L 的输入 $x(t)$ 与输出 $y(t)$ 以下式相联系时，

$$\sum_{j=0}^{n} a_j y^{(n-j)}(t)=\sum_{k=0}^{m} b_k x^{(m-k)}(t)$$

则在下列初始条件之下，

$$y^{(j)}(t_0)=0,\quad 0\leqslant j\leqslant n-1$$

L 也是一个线性时不变系统，而它的频率响应容易验证是

$$\Phi(\omega)=\sum_{j=0}^{m} b_j(i\omega)^j\Big/\sum_{j=0}^{n} a_j(i\omega)^j \tag{15}$$

对于输入为时间序列 x_n 的系统，也可规定其线性时不变性，以及系统的频率响应 $\Phi(\omega)$ 和脉冲响应 h_n，并且可以同样证明有如下的类似于(11)——(14)的公式

$$\Phi(\omega)=\sum_{n=-\infty}^{\infty} h_n e^{-in\omega} \tag{16}$$

$$Y(\omega)=\Phi(\omega)X(\omega),\ -\pi\leqslant\omega\leqslant\pi \tag{17}$$

$$y_n=\sum_{k=-\infty}^{\infty} h_{n-k}x_k \tag{18}$$

其中

$$X(\omega)=\sum_{n=-\infty}^{\infty} x_n e^{-in\omega}$$

称为 $\{x_n\}$ 的频谱，类似的，$Y(\omega)$ 为 $\{y_n\}$ 的频谱.

二、随机输入

当我们讨论系统的各种问题时，往往会遇到随机过程作为输入后的种种问题，例如湍流作用之下机翼的振动问题，热噪声对电

路输出的影响问题等等. 这时,对线性时不变系统中的平稳过程,我们也可以从频率域上来分析这些问题.

由于平稳过程的谱分解定理,对平稳过程 $X(t)$ 可看为下式的随机振动和 $X_n(t)$ 的极限

$$X_n(t)=\sum_j(\xi(\lambda_{j+1})-\xi(\lambda_j))e^{i\lambda_j t}$$

其中 $\xi(\lambda)$ 为 $X(t)$ 的随机谱函数. 因为 $X_n(t)$ 是随机振幅的简谐振动 $e^{i\lambda_j t}$ 之线性迭加,由(3)

$$\begin{aligned}Y_n(t)&=L[X_n(t)]\\&=\sum_j(\xi(\lambda_{j+1})-\xi(\lambda_j))L[e^{i\lambda_j t}]\\&=\sum_j\Phi(\lambda_j)(\xi(\lambda_{j+1})-\xi(\lambda_j))e^{i\lambda_j t}\end{aligned}$$

因此 $Y_n(t)$ 是一个平稳过程,而当 $X_n(t)$ 均方收敛于 $X(t)$ 时,若 L 又保持对均方收敛的连续性,就可得出对应随机输入 $X(t)$ 的输出 $Y(t)$ 为

$$\begin{aligned}Y(t)&=L[X(t)]=L[\lim X_n(t)]\\&=\lim L[X_n(t)]\\&=\lim\sum_j\Phi(\lambda_j)(\xi(\lambda_{j+1})-\xi(\lambda_j))e^{i\lambda_j t}\\&=\int_{-\infty}^{\infty}e^{i\lambda t}\Phi(\lambda)d\xi(\lambda)\end{aligned}$$

显然 $Y(t)$ 也是一个平稳过程,利用谱分解的唯一性,可以取 $Y(t)$ 的随机谱函数为

$$\xi_Y(\lambda)=\int_{-\infty}^{\lambda}\Phi(\mu)d\xi(\mu) \tag{19}$$

而 $Y(t)$ 的谱函数 $F_Y(\lambda)$ 与 $X(t)$ 的谱函数 $F_X(\lambda)$ 间必满足

$$F_Y(\lambda)=\int_{-\infty}^{\lambda}|\Phi(\mu)|^2dF_X(\mu)+C \tag{20}$$

特别当 $X(t)$ 存在谱密度 $f_X(\lambda)$ 时,则 $Y(t)$ 也必有如下谱密度

$$f_Y(\lambda)=|\Phi(\lambda)|^2 f_X(\lambda) \tag{21}$$

当输入 $X(t)$为平稳正态过程时，则其随机谱函数 $\xi(\lambda)$也必是正态过程，所以由(19)，$Y(t)$也必是一个正态过程，这样就有下列定理.

定理 1 若 $\Phi(\lambda)$对平稳过程 X_T 的谱函数 $F_X(\lambda)$满足

$$\int_{-\infty}^{\infty}|\Phi(\lambda)|^2 dF_X(\lambda)<\infty$$

又 L 是以 $\Phi(\lambda)$为频率响应的线性时不变系统，且关于输入的均方收敛为连续，则

(i) 当输入 $X(t)$为平稳过程时，输出 $Y(t)$也是平稳过程，且它们的谱函数以(20)相联系，而当 $X(t)$存在谱密度时，$Y(t)$也存在谱密度，且以(21)相联系；

(ii) 当输入 $X(t)$为平稳正态过程时，输出 $Y(t)$也是平稳正态过程.

对输入输出为时间序列的线性时不变系统来说，类似的结论也是成立的，我们不再重复了.

[例1] 考虑图 1 中的系统.若以 $x(t)$和 $y(t)$ 分别表示输入和输出的电压，则它们可用如下线性微分方程相联系

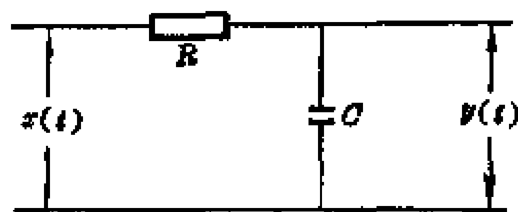


图 1

$$\frac{dy(t)}{dt}+\alpha y(t)=\alpha x(t)$$

这里 $\alpha=\frac{1}{RC}$. 所以它是一个线性时不变系统. 而且按(15)，它的频率响应为

$$\Phi(\lambda)=\frac{\alpha}{i\lambda+\alpha}, \qquad -\infty<\lambda<\infty$$

我们考虑输入为平稳过程 $X(t)$的情况，若 $X(t)$的协方差函数

$$B_X(\tau)=\sigma_0^2 e^{-\beta|\tau|},\quad \beta>0$$

则按(2.6)可得它的谱密度为

$$f_X(\lambda)=\frac{2}{\pi}\sigma_0^2\frac{\beta}{\beta^2+\lambda^2}$$

所以由定理 1 输出 $Y(t)$ 为平稳过程，其谱密度为

$$f_Y(\lambda)=|\Phi(\lambda)|^2 f_X(\lambda)$$

$$=\frac{2}{\pi}\sigma_0^2\frac{\alpha^2\beta}{(\alpha^2+\lambda^2)(\beta^2+\lambda^2)}$$

进而可算出 $Y(t)$ 的协方差函数为

$$B_Y(\tau)=\frac{2}{\pi}\int_0^\infty f_Y(\lambda)\cos\tau\lambda d\lambda$$

$$=\frac{2}{\pi}\sigma_0^2\alpha^2\beta\int_0^\infty\frac{\cos\tau\lambda}{(\alpha^2+\lambda^2)(\beta^2+\lambda^2)}d\lambda$$

$$=\frac{\alpha\sigma_0^2}{\alpha^2-\beta^2}(\alpha e^{-\beta\tau}-\beta e^{-\alpha\tau}),\quad \tau\geqslant 0$$

特别
$$D(Y)=B_Y(0)=\frac{\alpha}{\alpha+\beta}\sigma_0^2$$

所以利用(21)可以估计输出信号的统计特征. 当 $X(t)$ 为正态过程时，我们更可由此定出 $Y(t)$ 的有限维分布，也可计算出 $Y(t)$ 在一定范围内取值的概率.

[例 2] 在利用数字计算机对波形信号进行处理时，经常要进行如下的运算

$$y(n)=-\sum_{j=1}^{p}a_j y(n-j)+\sum_{j=0}^{q}b_j x(n-j) \tag{22}$$

其中 $\{x(n)\}$ 为被处理的信号，$\{y(n)\}$ 表示处理以后的信号，其中 a_j, b_j 为已知的常数，通常可保证 $A(z)=z^p+\sum_{j=1}^{p}a_j z^{p-j}$ 的零点全在圆内，这就是所谓递归滤波. 由于计算机存放每个数据的字长有限，所以往往对输入及中间运算结果进行舍入，只保留有限位，

这样实际上的输出是

$$\hat{y}(n) = -\sum_{j=1}^{p} a_j \hat{y}(n-j) + \sum_{j=0}^{q} b_j x(n-j) + \varepsilon(n)$$

其中 $\varepsilon(n)$ 是由于对输入和中间运算结果进行舍入而引起的误差把它与(22)式相减，可得输出误差 $\tilde{y}(n) \triangleq \hat{y}(n) - y(n)$ 满足

$$\tilde{y}(n) = -\sum_{j=1}^{p} a_j \tilde{y}(n-j) + \varepsilon(n) \tag{23}$$

若为定点运算，那么可认为 $\{\varepsilon(n)\}$ 是相互独立同分布的随机变量序列，且 $\varepsilon(n)$ 都是 $[-C2^{-b}, C2^{-b}]$ 上均匀分布的，其中 b 表定点运算的字长，C 为某常数①，为了考虑输出误差的统计特性，可考虑(23)的稳态情况，这时 $\{\tilde{y}(n)\}$ 就是以 $\{\varepsilon(n)\}$ 作为平稳输入时的平稳输出，而这一系统的频率特征由(23)就是

$$\Phi(\lambda) = \frac{1}{1 + \sum_{j=1}^{p} a_j e^{-ij\lambda}}$$

由于 $\{\varepsilon(n)\}$ 为白噪声

$$D(\varepsilon(n)) = \frac{C^2}{6} 2^{-2b}$$

所以由(21)，$\tilde{y}(n)$ 的谱密度是

$$f_{\tilde{y}}(\lambda) = \frac{1}{2\pi \left| 1 + \sum_{j=1}^{p} a_j e^{-ij\lambda} \right|^2} \cdot \frac{C^2}{6} 2^{-2b}$$

输出误差 $\{\tilde{y}(n)\}$ 的方差就是

---

① 参见 L. R. Rabiner, B. Gold, "Theory and Applicalion of Digital Signal Processing"(1975).

$$D(\tilde{y}(n))=\frac{C^2}{12\pi}2^{-2b}\int_{-\pi}^{\pi}\frac{d\lambda}{|1-\sum_{j=1}^{p}a_je^{-ij\lambda}|^2}$$

这就是在信号数字处理中递归滤波估计有限字长效应的一种最简单的模型和做法.

三、平稳相关过程与互谱函数

当我们讨论系统的随机输入输出时，涉及的已不仅是一个平稳过程，而是输入 $X(t)$ 和输出 $Y(t)$，对于两个或多个随机过程，除了各别研究这些过程外，还需研究它们的相互关系，这时平稳性的概念除了要求每个过程本身是平稳的以外，还要求它们之间的统计联系是平稳的.

定义 平稳过程 $X_T=\{X(t), -\infty<t<\infty\}$ 和平稳过程 $Y_T=\{Y(t), -\infty<t<\infty\}$ 称为**平稳相关**的，若对任何的 h 成立等式

$$E\{X(t+h)\overline{Y(s+h)}\}=E\{X(t)\overline{Y(s)}\}$$

由(23)我们可以如下规定 X_T 和 Y_T 的**互协方差函数**:

$$B_{XY}(t)=E\{(X(t+s)-EX(t+s))\overline{(Y(s)-EY(s))}\} \quad (24)$$

并且不难推出

$$B_{XY}(t)=\overline{B_{YX}(-t)}$$

若 X_T, Y_T 是平稳相关的，那么对任何常数 C, $X(t)+CY(t)$ 也是平稳过程，且它的协方差函数 $B_{X+CY}(t)$ 满足

$$B_{X+CY}(t)=B_X(t)+\bar{C}B_{XY}(t)+CB_{YX}(-t)+|C|^2B_Y(t) \quad (25)$$

利用(25)可以推出 X_T, Y_T 的互协方差函数 $B_{XY}(t)$ 也有谱分解式，这只要让(25)中的 C 取 $+1, -1, +i, -i$，立即可推出

$$B_{XY}(t)=\frac{1}{4}[B_{X+Y}(t)-B_{X-Y}(t)+iB_{X+iY}(t)-iB_{X-iY}(t)]$$

若以 $F_{X+CY}(\lambda)$ 表过程 $X(t)+CY(t)$ 的谱函数，又令

$$F_{XY}(\lambda)=\frac{1}{4}[F_{X+Y}(\lambda)-F_{X-Y}(\lambda)+iF_{X+iY}(\lambda)-iF_{X-iY}(\lambda)]$$

则对 $B_{XY}(t)$ 有如下谱分解式

$$B_{XY}(t)=\int_{-\infty}^{\infty}e^{it\lambda}dF_{XY}(\lambda) \tag{26}$$

特别当 $F_{XY}(\lambda)$ 为绝对连续时，更可表为

$$B_{XY}(t)=\int_{-\infty}^{\infty}e^{it\lambda}f_{XY}(\lambda)d\lambda$$

其中 $F_{XY}(\lambda)$, $f_{XY}(\lambda)$ 又分别称为 X_T, Y_T 的**互谱函数**和**互谱密度**，当然 $f_{XY}(\lambda)$ 不一定是非负的，而 $F_{XY}(\lambda)$ 只是有界变差的，不一定是单调不减的。利用(25)式及上面的方法也可以说明 $F_{XY}(\lambda)$ 亦由 $B_{XY}(t)$ 唯一确定(不计常数之差)。

对时不变系统，若 X_T 为输入的零均值平稳过程，而 Y_T 为对应于 X_T 的输出，则由(19)，

$$B_{YX}(t)=E\{Y(t+s)\overline{X(s)}\}=\int_{-\infty}^{\infty}e^{it\lambda}\Phi(\lambda)dF_X(\lambda)$$

所以 Y_T, X_T 的互谱函数 $F_{YX}(\lambda)$ 可取为

$$F_{YX}(\lambda)=\int_{-\infty}^{\lambda}\Phi(\mu)dF_X(\mu)$$

特别当 X_T 存在谱密度 $f_X(\lambda)$ 时，则 Y_T, X_T 也有互谱密度

$$f_{YX}(\lambda)=\Phi(\lambda)f_X(\lambda) \tag{27}$$

(27)和(21)同样是线性时不变系统中的一个重要公式，(21)帮助我们分析输入输出的统计特征，而(27)可以从输入功率谱和输出输入互谱中来求得系统的频率特征，作为分析改进系统的依据。在船舶的适航性研究中，就可从海浪波高(输入)和相应的船舶摇摆(输出)的记录中获得船舶(系统)的频率特性，在汽车振动的研究和飞机受湍流影响的研究中也有类似的情况，这种试验可以是试

航中的实际记录，也可以是用模型在水池，风洞等实验设备中进行，当然得到了实测记录后首先要由实测记录估计出输入信号的功率谱和输出入的互谱，这一点我们将在下一章中介绍，而有了 $f_X(\lambda)$ 和 $f_{YX}(\lambda)$ 的估计后，(27) 就是由此求得系统频率特性 $\Phi(\lambda)$的依据了.

四、均方遍历定理

在概率论基础中，我们已经接触过大数法则，即对独立同分布随机变量序列$\{X(n), n\geq 1\}$，若存在均值 $EX(n)=m$，则

$$\lim_{N\to\infty} P\left(\left|\frac{1}{N}\sum_{n=1}^{N} X(n)-m\right|>\varepsilon\right)=0 \tag{28}$$

用这一篇的概念来看，$\{X(n),n\geq 1\}$ 也是一个随机过程，$\frac{1}{N}\sum_{n=1}^{N} X(n)$ 是对随机过程的样本按时间取平均，它随不同样本取不同数值，也是一个随机变量，而 $m=EX(n)$ 是随机过程的均值，即在某一时刻随机过程的现实取值的统计平均 $m=EX(n)$，所以(28) 表明随着时间的增长，随机过程样本按时间的平均值以越来越大的概率无限接近于随机过程的统计平均．也就是说，对于这个随机过程，只要我们进行观察的时间足够长，它每个样本都能“遍历”各种可能的状态，因而一个样本按时间的平均就可以近似也代替它在固定时刻取值的统计平均．随机过程的这个性质也称为**遍历性**，当然，这只有在不同时刻的统计平均是相等的这一前提下才有可能．所以我们对平稳过程来考虑它的遍历性.

定理 2 若 $X_T=\{X(t), -\infty<t<\infty\}$为均方连续平稳过程，$EX(t)=m$，则

$$\lim_{T\to\infty} E\left|\frac{1}{2T}\int_{-T}^{T} X(t)dt-m\right|^2=0 \tag{29}$$

的充要条件是 X_T 的谱函数 $F(\lambda)$在 $\lambda=0$ 连续或者等价地，X_T 的

协方差函数 $B(t)$满足下列条件

$$\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}B(t)dt=0 \tag{30}$$

［证明］ 不妨假设 $m=0$,否则令 $Y(t)=X(t)-m$ 改为考虑 $Y(t)$即可. 由于 X_T 均方连续,所以有谱分解式

$$X(t)=\int_{-\infty}^{\infty}e^{it\lambda}d\zeta(\lambda)$$

而且对 $X(t)$进行均方积分时,由定理 3.2.16 有

$$\frac{1}{2T}\int_{-T}^{T}X(t)dt=\int_{-\infty}^{\infty}\left(\frac{1}{2T}\int_{-T}^{T}e^{it\lambda}dt\right)d\zeta(\lambda)$$
$$=\int_{-\infty}^{\infty}\Phi_T(\lambda)d\zeta(\lambda)$$

其中

$$\Phi_T(\lambda)=\begin{cases}\dfrac{\sin T\lambda}{T\lambda}, & \lambda\neq 0\\ 1, & \lambda=0\end{cases}$$

由(2.40)可得

$$E\left|\frac{1}{2T}\int_{-T}^{T}X(t)dt\right|^2=\int_{-\infty}^{\infty}\Phi_T^2(\lambda)dF(\lambda)$$

因为

$$|\Phi_T(\lambda)|\leqslant 1$$

$$\lim_{T\to\infty}\Phi_T(\lambda)=\begin{cases}0, & \lambda\neq 0\\ 1, & \lambda=0\end{cases}$$

所以利用 $\Phi_T(\lambda)$的有界收敛性可得

$$\lim_{T\to\infty}E\left|\frac{1}{2T}\int_{-T}^{T}X(t)dt\right|^2$$
$$=\lim_{T\to\infty}\int_{-\infty}^{\infty}\Phi_T^2(\lambda)dF(\lambda)$$
$$=F(0+)-F(0)$$

因此(29)成立的充要条件是 $F(\lambda)$在 $\lambda=0$ 连续. 另外, 利用协方

差函数的谱分解式可得

$$\frac{1}{2T}\int_{-T}^{T}B(t)dt$$

$$=\int_{-\infty}^{\infty}\left(\frac{1}{2T}\int_{-T}^{T}e^{it\lambda}dt\right)dF(\lambda)$$

$$=\int_{-\infty}^{\infty}\Phi_T(\lambda)dF(\lambda)$$

所以

$$\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}B(t)dt$$

$$=\lim_{T\to\infty}\int_{-\infty}^{\infty}\Phi_T(\lambda)dF(\lambda)$$

$$=F(0+)-F(0)$$

即 $F(\lambda)$ 在 $\lambda=0$ 连续与(30)是等价的，定理证毕。

在实用中，可以用$\frac{1}{T}\int_0^T X(t)dt$ 代替(29)中的$\frac{1}{2T}\int_{-T}^{T}X(t)dt$，定理的结论不变，此外对离散参数的平稳序列，也有类似的结论。

对于实的平稳正态过程来说，由于它的有限维分布对时间推移是不变的，所以 $\{X^2(t),\ -\infty<t<\infty\}$，$\{X(t+\tau)X(t),\ -\infty<t<\infty\}$都是平稳过程．若令

$$Y(t)=X^2(t)-B_X(0)$$

则

$$EY(t)=0$$

$$B_Y(\tau)\triangleq EY(t+\tau)Y(t)=2B_X^2(\tau)$$

若记 $F(\lambda)$为 X_T 的谱函数，则

$$\frac{1}{2T}\int_{-T}^{T}B_X{}^2(\tau)d\tau$$

$$=\frac{1}{2T}\int_{-T}^{T}\left(\iint_{-\infty}^{\infty}e^{i\tau(\lambda-\mu)}dF(\lambda)dF(\mu)\right)d\tau$$

$$=\iint_{-\infty}^{\infty} D_T(\lambda-\mu)dF(\lambda)dF(\mu)$$

其中

$$D_T(\lambda-\mu)=\begin{cases}\dfrac{\sin T(\lambda-\mu)}{T(\lambda-\mu)}, & \lambda\neq\mu\\ 1, & \lambda=\mu\end{cases}$$

所以

$$\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T} B_X{}^2(\tau)d\tau$$

$$=\lim_{T\to\infty}\iint_{-\infty}^{\infty} D_T(\lambda-\mu)dF(\lambda)dF(\mu)$$

$$=\sum_j(\Delta F(\lambda_j))^2$$

其中$\{\lambda_j\}$为F的跳跃点全体，$\Delta F(\lambda_j)$表示F在λ_j的跃度，所以由定理 2, $Y(t)$均方遍历的充要条件是X_T的谱函数连续. 当F连续时，同样可以证明

$$Y_2(t)=X(t)X(t+\tau)-B(\tau)$$

也是均方遍历的，所以我们可以通过

$$\frac{1}{T}\int_0^{T-|\tau|} X(t)X(t+\tau)dt-\left(\frac{1}{T}\int_0^T X(t)dt\right)^2$$

来估计$B_X(\tau)$，这将是第五章中讨论平稳过程各种参数估计问题的出发点.

五、平稳过程的采样定理

在运用计算机对随时间连续变化的信号$x(t)$进行处理时，首先必须对信号进行离散的"采样"(注意这里使用了"采样"这个词，它与统计中的随机取样不是同一个概念)，一般是在每隔定长的时间Δ后对$x(t)$进行观察，获得$x(t)$在$t=k\Delta$时的数值(见图 2)

$$x_k=x(k\Delta)$$

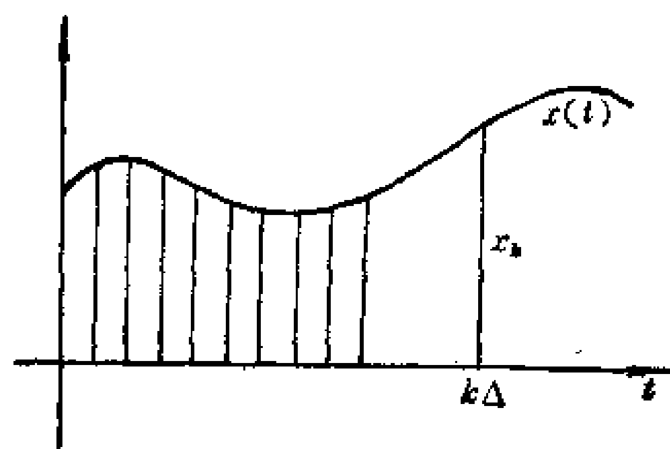


图 2

这个过程可由专用的设备来完成，但相隔多长的时间Δ进行采样呢，这是一个值得讨论的问题，过密的采样会加重采样设备的负担，也增加处理运算的工作量，但过稀采样将使采样值无法复原，例如在图3上不同频率的正弦波有相同的采样值，这样从获得的采样值就无法辨认它是哪一个波形的采样值，对它的处理也不会合理，

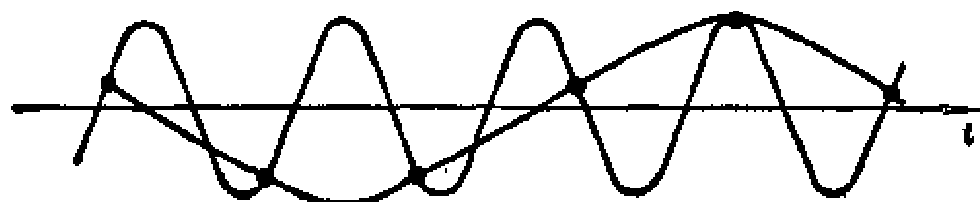


图 3

处理的结果也难于恢复成正确的连续波形．对于非随机的决定性的信号，采样定理回答了这一问题，即当波形频谱中最高频率小于B时，那么$\Delta<\frac{1}{2B}$时，在理论上可以由采样值$\{x(k\Delta)\}$完全准确地恢复$x(t)$．对于平稳过程，也有类似的结果．

定理3 若$X_T=\{X(t),\ -\infty<t<\infty\}$的均方连续平稳随机过程，又$X_T$的谱函数$F(\lambda)$满足

$$\int_{|\lambda|\geq 2\pi B} dF(\lambda)=0 \tag{31}$$

则取

$$\Delta=\frac{1}{2B} \tag{32}$$

便有

$$X(t)=\sum_{k=-\infty}^{\infty} X(k\Delta)\frac{\sin\frac{\pi}{\Delta}(t-k\Delta)}{\frac{\pi}{\Delta}(t-k\Delta)} \tag{33}$$

上述级数对每个 t 按均方意义下是收敛的.

[证明]由于(31), $X(t)$的谱分解可写为

$$X(t)=\int_{|\lambda|<2\pi B} e^{i\lambda t}d\xi(\lambda)$$

对固定的 t, 在$[-2\pi B, 2\pi B]$上我们考虑函数 $e^{i\lambda t}$, 若将 $e^{i\lambda t}$ 展为如下富利埃级数

$$e^{i\lambda t}\sim\sum_{n=-\infty}^{\infty} a_n e^{in\frac{\lambda}{2B}} \tag{34}$$

则

$$\begin{aligned}a_n&=\frac{1}{4\pi B}\int_{-2\pi B}^{2\pi B} e^{i\lambda t}e^{-in\frac{\lambda}{2B}}d\lambda\\&=\frac{\sin\left[\left(t-\frac{n}{2B}\right)\cdot 2\pi B\right]}{2\pi B\left(t-\frac{n}{2B}\right)}\\&=\frac{\sin\frac{\pi}{\Delta}(t-n\Delta)}{\frac{\pi}{\Delta}(t-n\Delta)}\end{aligned}$$

由于 $e^{i\lambda t}$ 是 λ 的连续函数, 且在$[-2\pi B, 2\pi B]$上有界变差, 则由富利埃级数理论可以知道①, (34)中级数的部分和一致有界, 且在$(-2\pi B, 2\pi B)$上收敛于 $e^{i\lambda t}$, 所以有

---

① 参见菲赫金哥尔茨"微积分学教程"第三卷第三分册.

$$\lim_{N\to\infty}\int_{|\lambda|<2\pi B}\left|e^{i\lambda t}-\sum_{n=-N}^{N}a_n e^{i\pi\frac{\lambda}{2B}}\right|^2 dF(\lambda)=0$$

利用§2中提到 $H_T(X)$ 与 $L_2(dF)$ 的对应，便有

$$\lim_{N\to\infty}E\left|X(t)-\sum_{n=-N}^{N}a_n X\left(\frac{n}{2B}\right)\right|^2=0$$

所以在均方收敛的意义下，$X(t)$ 可表为

$$X(t)=\sum_{n=-\infty}^{\infty}a_n X\left(\frac{n}{2B}\right)$$

$$=\sum_{n=-\infty}^{\infty}X(n\Delta)\frac{\sin\frac{\pi}{\Delta}(t-n\Delta)}{\frac{\pi}{\Delta}(t-n\Delta)}$$

这就是定理的结论.

上述定理也称为平稳过程的**采样定理**，实用中常遇到的情况是 X_T 有谱密度 $f(\lambda)$ 的情况，这时相应于(31)的对 X_T 的要求是

$$f(\lambda)=0,\quad |\lambda|\geqslant\frac{1}{2B}\tag{35}$$

采样定理表明，(31)或(35)成立时，$X(t)$ 可由它的采样值 $X(k\Delta)$ 按(33)计算出来，但在实用中，只能获得 $X(k\Delta)$ 的有限个采样值. 若用它们来代表 $X(t)$ 的波形时，所使用的采样率就应比(32)所规定的要高.

利用了采样定理的结果，对时间为连续的信号，不论是决定性或是随机的，都可根据其谱的性质在一定采样率之下进行采样取值，这样做并不损失信号的任何信息，而且也只有取其采样值才能在数字计算机上处理与运算. 对于平稳过程 $X_T=\{X(t),\ -\infty<t<\infty\}$ 来说，它的采样值 $X=\{X(n\Delta),\ n=0,\ \pm1,\cdots\}$ 就是一个平稳时间序列.

六、白噪声

最后，我们来介绍在电路系统分析，自动控制和测量中经常遇到的一类随机干扰——“白噪声”．因为在电路系统中，由于分子的热运动，使电路各处的电流或电压受到随机干扰，在系统分析中，也把随机干扰称为噪声，因为当这种电压或电流的变化可反映为声波的变化时，就是一种人们不爱听的嘶嘶嚓嚓的噪声．从数学上看，这就是随机过程．当外界条件基本不变时，又可认为系统中或测量中遇到的噪声的主要统计特性是不随时间推移而改变的，所以它还是平稳过程．从功率谱角度来看，这种噪声对不同频率的输入都能进行干扰，所以它的谱在各个频率分量上都广泛地存在着．一种常用的抽象是把这种噪声假定为在各个频率分量上都有同样的功率．类似于白光的能谱在各种频率上均匀分布，我们也把这类噪声称为**白噪声**，所以白噪声是功率谱密度为常数的零均值平稳过程．

在参数为离散的情况，若 $X_T=\{X(n),\ n=0,\ \pm 1,\cdots\}$ 为白噪声，其功率谱密度

$$f(\lambda)=\frac{\sigma^2}{2\pi}$$

则

$$EX(n)=0$$

$$E\{X(n)\overline{X(m)}\}=\int_{-\pi}^{\pi} e^{i(n-m)\lambda}\,\frac{\sigma^2}{2\pi}d\lambda$$

$$=\begin{cases}\sigma^2, & m=n\\ 0, & m\neq n\end{cases}$$

所以 X_T 为具有相同方差且互不相关的随机变量序列，这是在离散时间序列分析中描述随机干扰经常采用的一种模型，也是较为简单的一种，特别当 X_T 还是正态过程时，它就是相互独立同分布正态随机变量序列．

在参数为连续的情况下，若 $X_T=\{X(t),\ -\infty<t<\infty\}$ 的功

率谱密度 $f(\lambda)=\sigma^2$，由于

$$E|X(t)|^2=B(0)$$
$$=\int_{-\infty}^{\infty}f(\lambda)d\lambda=+\infty$$

所以 X_T 已不是通常意义下的平稳过程，因为它不存在方差，这时上面规定的白噪声就是一种数学上的抽象，它可以看为现实噪声的一种理想化．例如考虑平稳过程 $X_T^{(\rho)}$，它的谱密度 $f_\rho(\lambda)$ 和协方差函数 $B_\rho(\tau)$ 分别是

$$f_\rho(\lambda)=\frac{\sigma^2}{\pi}\frac{1}{1+(\lambda/\rho)^2} \tag{36}$$

$$B_\rho(\tau)=\rho\sigma^2e^{-\rho|\tau|},\ \rho>0 \tag{37}$$

则当 $\rho\to\infty$ 时 $X_T^{(\rho)}$ 的极限状态 X_T 就设想它是一个白噪声，从(36)看，$f_\rho(\lambda)$ 的极限是常数 $\frac{\sigma^2}{\pi}$，从(37)看，$B_\rho(\tau)$ 的极限是一个 δ 函数(不计常数因子差别)，即 X_T 的方差是无限的，而 $X(t)$ 和 $X(t')$ 当 $t\neq t'$ 时是互不相关的，如果我们形式的写出白噪声的谱分解式

$$X(t)=\int_{-\infty}^{\infty}e^{it\lambda}d\zeta(\lambda)$$

并对 $X(t)$ 积分，令

$$Y(t)=\int_0^t X(t)dt$$
$$=\int_{-\infty}^{\infty}\frac{e^{it\lambda}-1}{i\lambda}d\zeta(\lambda) \tag{38}$$

只要 $\zeta(\lambda)$ 是满足

$$E|\zeta(\lambda_2)-\zeta(\lambda_1)|^2=\frac{\sigma^2}{2\pi}|\lambda_2-\lambda_1|$$

的正交增量过程，(38)右端的随机积分是有完全确定意义的，且由富利埃积分的巴塞伐尔等式，可以得到

$$\int_{-\infty}^{\infty}\frac{e^{it_2\lambda}-e^{it_1\lambda}}{i\lambda}\cdot\frac{e^{-it_2\lambda}-e^{-it_1\lambda}}{-i\lambda}d\lambda$$

$$=2\pi\int_{-\infty}^{\infty}\chi_{(t_1,t_2)}(t)\chi_{(t_3,t_4)}(t)dt$$

这样就有

$$E\left\{\int_{-\infty}^{\infty}\frac{e^{it_2\lambda}-e^{it_1\lambda}}{i\lambda}d\zeta(\lambda)\cdot\overline{\int_{-\infty}^{\infty}\frac{e^{it_4\lambda}-e^{it_3\lambda}}{i\lambda}d\zeta(\lambda)}\right\}$$

$$=\sigma^2\int_{-\infty}^{\infty}\chi_{(t_1,t_2)}(t)\chi_{(t_3,t_4)}(t)dt \tag{39}$$

其中 $\chi_A(t)$是集 A 的示性函数，所以当(t_1, t_2)，(t_3, t_4)互不相交时，

$$E\left\{\int_{t_1}^{t_2}X(t)dt\overline{\int_{t_3}^{t_4}X(t)dt}\right\}=0$$

因此 $Y_T=\{Y(t), -\infty<t<\infty\}$是一个正交增量过程，且由(39)可知

$$E|Y(t)|^2=\sigma^2 t \tag{40}$$

这样，白噪声 X_T 还可形式地理解为满足(40) 的正交增量过程 Y_T 的均方导数，即

$$X(t)=Y'(t)$$

当 X_T 还是正态过程，Y_T 就是满足(40)的独立增量过程，即维纳过程或布朗运动，这时 X_T 就是布朗运动的形式上的导数。

§4. 自回归滑动和过程

在这一节里，我们将考虑满足下列方程的随机过程 $X_T=\{X(n), n=0, \pm1, \cdots\}$，

$$X(n)+a_1X(n-1)+\cdots+a_pX(n-p)$$
$$=b_0\xi(n)+b_1\xi(n-1)+\cdots+b_q\xi(n-q) \tag{1}$$

其中$\{\xi(n), n=0, \pm1, \cdots\}$为白噪声或其它随机过程，这是在自动控制等方面经常遇到的一类过程，我们要对它进行讨论是出于以

下一些原因:

首先在近代控制理论中引入了"状态"的概念，从而系统的变化发展都可由它在某个时刻的"状态"和它的运动方程来描述．例如匀速运动的飞机就可把它的位置坐标 $x(t)$，$y(t)$，$z(t)$和速度 $\dot{x}(t)$，$\dot{y}(t)$，$\dot{z}(t)$作为状态，而它的运动方程就是：

$$\frac{d}{dt}\begin{bmatrix} x(t) \\ y(t) \\ z(t) \\ \dot{x}(t) \\ \dot{y}(t) \\ \dot{z}(t) \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ & & & 0 & & \\ & & & & & \\ & & & & & \end{bmatrix} \begin{bmatrix} x(t) \\ y(t) \\ z(t) \\ \dot{x}(t) \\ \dot{y}(t) \\ \dot{z}(t) \end{bmatrix}$$

系统的状态在一般情况下是一个多维向量 $Y(t)$，假若系统又有随机的干扰或输入，那么经常遇到的是下列形式的方程

$$\frac{d}{dt}Y(t) = \Phi(t)Y(t) + \eta(t) \tag{2}$$

其中 $\Phi(t)$为一矩阵，$\eta(t)$代表随机输入或干扰，这时 $Y(t)$就是一个随机过程，它的离散形式是

$$Y(n+1) = F(n)Y(n) + \eta(n) \tag{3}$$

$Y(n)$是一个随机向量，假若考虑它的某个分量 $Y_k(n)$，那么以后将证明它与(1)中的随机过程有密切的联系．

其次，假若我们考虑如下线性时不变系统

$$\begin{aligned} & y(n) + a_1 y(n-1) + \cdots + a_p y(n-p) \\ & = b_0 x(n) + b_1 x(n-1) + \cdots + b_q x(n-q) \end{aligned} \tag{4}$$

其中 $x(n)$表示输入，$y(n)$表示输出，那么(1)规定的 X_T 可看为上述系统在输入为 ξ_T 时的输出，而形为(4)的系统又是最常见的一种离散的线性时不变系统，所以(1)中的 X_T 也是经常遇见的．

此外，在上两节可以看到，一个平稳过程可以用它的协方差函数或者等价地用一个谱函数来刻划它的主要统计特征，但是协方

差函数或谱都要由无限个数值来描述．所以在给出一个平稳过程或由实际资料统计出平稳过程的主要统计特征上都还感到不方便．若 ξ_T 是白噪声，则对于满足(1)的平稳过程类，往往只需要有限个参数 a_j, b_k 等就可定出它的谱或协方差函数，而由协方差函数或谱在有限个点上的数值，一般也可定出(1)中的 a_j，b_k，从而规定出 X_T 满足的方程式(1)，所以满足(1)的平稳过程是一种有限参数模型．并且由于它满足(1)，在用于预测时有不少好处，在数学处理上也较为方便，所以(1)就成为平稳过程的一种重要的有限参数模型作为我们近似描述一般平稳过程的工具．

对于(1)中的平稳过程，当 $p=0$ 时，$X(n)$ 可表为

$$X(n)=b_0\xi(n)+b_1\xi(n-1)+\cdots+b_q\xi(n-q) \tag{5}$$

这时我们称 X_T 是 ξ_T 的**滑动和**，或称 X_T 满足滑动和模型，用 MA 表示．有时为了强调(5)的右端是有限项之和，又称(5)为**有限滑动和**，(5)中的 q 又称为滑动和的**阶**．

若在(1)中，$q=0$，则 $X(n)$ 可表为

$$X(n)+a_1X(n-1)+\cdots+a_pX(n-p)=b_0\xi(n) \tag{6}$$

或

$$X(n)=-a_0X(n-1)-\cdots-a_pX(n-p)+b_0\xi(n) \tag{7}$$

在(7)中 $X(n)$ 表为其自身以前的值 $X(n-1)$，…，$X(n-p)$ 的线性回归．所以满足(6)，(7)的 $X(n)$ 又称为**自回归过程**或 X_T 满足自回归模型，或用 AR 表示，而(6)中的 p 又称为自回归的**阶**．

一般地，(1)中的 X_T 就称为**自回归滑动和过程**，或称 X_T 满足**自回归滑动和混合模型**，以 ARMA 表示．在这一节里，我们将从数学上来讨论 ARMA 过程的性质，为了叙述方便，在这一节里都假定所考虑的过程均值为零．在以后的讨论中，我们还要用到具有二阶矩的随机变量希尔伯脱空间 H 的如下子空间：

$H_T(X)$：由 $\{X(t), t\in T\}$ 张成的子空间

$H_n(X)$：由$\{X(t), t\leqslant n\}$张成的子空间

所以$H_T(X)$就是由$X(t)$，$t\in T$的有限线性组合及其均方极限全体构成的. 此外，对随机变量X提到它的内积和范数时都是指H中的内积和范数，即$(X, Y)=EX\overline{Y}$，$\|X\|=(E|X|^2)^{\frac{1}{2}}$.

引理1 若随机过程$X(n)$，$W(n)$满足如下方程

$$\sum_{j=0}^{p}a_jX(n-j)=W(n), \quad a_0=1 \tag{8}$$

则必存在常数c_l和$b_j(n,k)$，使$X(n)$可表为

$$X(n)=\sum_{l=0}^{k}c_lW(n-l)+\sum_{j=1}^{p}b_j(n,k)X(n-k-j) \tag{9}$$

特别当

$$\sum_{j=0}^{p}a_j\lambda^{p-j}=0 \tag{10}$$

的根的模全小于1，且

$$E|X(n)|^2\leqslant M, \quad E|W(n)|^2\leqslant M \tag{11}$$

则$X(n)$可表为

$$X(n)=\sum_{l=0}^{\infty}c_lW(n-l) \tag{12}$$

且$\sum\limits_{l=0}^{\infty}|c_l|<\infty$.

[证明] 为了证明(9)，只要对k运用数学归纳法即可. 此外，利用线性差分方程的知识①或母函数方法，可以得到若$X(n)$满足(8)，则它的解必可写为(9)，且当(10)无重根时，其中的c_l，$b_j(n,k)$必有形式

$$c_l=\sum_{s=1}^{p}h_s\lambda_s^l$$

① 参见福田武雄，差分方程，穆宏基译，上海科技出版社，1962.

$$b_j(n,k)=\sum_{s=1}^{p} b_{js}(n)\lambda_s^k$$

其中 λ_s, $1\leqslant s\leqslant p$ 为(10)的根，h_s, $b_{js}(n)$ 为常数. 所以当

$$|\lambda_s|<1,\quad 1\leqslant s\leqslant p$$

就有

$$\lim_{k\to\infty} b_j(n,k)=0 \tag{13}$$

$$\sum_{l=0}^{\infty}|c_l|<\infty \tag{14}$$

因此由(11)可得

$$\left\|\sum_{j=1}^{p} b_j(n,k)X(n-k-j)\right\|\leqslant\sum_{j=1}^{p}|b_j(n,k)|\cdot M$$

$$\left\|\sum_{l=N}^{N+q} c_l W(n-l)\right\|\leqslant\sum_{l=N}^{N+q}|c_l|\cdot M \tag{15}$$

所以由(13), (14),

$$\begin{aligned}X(n)&=\lim_{k\to\infty}\sum_{l=0}^{k} c_l W(n-l)\\&\quad+\lim_{k\to\infty}\sum_{j=1}^{p} b_j(n,k)X(n-k-j)\\&=\sum_{l=0}^{\infty} c_l W(n-l)\end{aligned}$$

由(15), 上述级数是均方收敛的, 这是当(10)无重根的情形, 当(10)有重根时, 证明也可同样的进行.

现在我们来讨论有限滑动和过程的一些性质.

定理 1 若 $X_T=\{X(n), n=0, \pm1, \cdots\}$ 为平稳随机序列, 则下列三个条件等价:

(*a*) $X(n)$ 可表为如下的 q 阶滑动和

$$X(n)=\sum_{k=0}^{q} b_k\xi(n-k) \tag{16}$$

其中$\{\xi(n)\}$为白噪声;

(b) X_T 的协方差函数 $B(n)$满足下列条件

$$B(n)=0, \quad |n|>q \tag{17}$$

(c) X_T 存在如下形式的谱密度 $f(\lambda)$,

$$f(\lambda)=\frac{\sigma^2}{2\pi}\left|\sum_{k=0}^{q}b_k e^{-ik\lambda}\right|^2 \tag{18}$$

且当 X_T 满足上述条件时,则 $X(n)$必可表为

$$X(n)=\sum_{k=0}^{q}\tilde{b}_k\tilde{\xi}(n-k)$$

其中$\{\tilde{\xi}(n)\}$为白噪声,而

$$\sum_{k=0}^{q}\tilde{b}_k z^{q-k}=0$$

的根的模都不超过1,并且

$$\tilde{\xi}(n)\in H_n(X) \tag{19}$$

[证明] $(a)\to(b)$: 直接利用(16)计算 $X(n)$ 的协方差函数 $B(n)$即可得(17)。

$(b)\to(c)$: 由(17)可知 X_T 存在如下的谱密度

$$f(\lambda)=\frac{1}{2\pi}\sum_{k=-q}^{q}B(k)e^{-ik\lambda}$$

若记

$$A(z)=\frac{1}{2\pi}\sum_{k=-q}^{q}B(k)z^{-k}$$

$$A^*(z)=\frac{1}{2\pi}\sum_{k=-q}^{q}\overline{B(k)}z^{-k}$$

则因为 $f(\lambda)$是实的, $f(\lambda)=\overline{f(\lambda)}$,

$$A(e^{i\lambda})=f(\lambda)=\overline{f(\lambda)}=\overline{A(e^{i\lambda})}=A^*\left(\frac{1}{e^{i\lambda}}\right)$$

所以利用解析函数的唯一性定理可得

$$A(z)=A^*\left(\frac{1}{z}\right)$$

因此，$A(z)$的根必关于圆成对共轭地出现，即若 β_i 为 $A(z)$的根，必有$\dfrac{1}{\bar{\beta}_i}$也是它的根，则可取 $A(z)$在$|z|\leqslant 1$ 的根为$\{\beta_i, 1\leqslant i\leqslant q\}$，并把 $A(z)$表为

$$A(z)=cz^{-q}\prod_{j=1}^{q}(z-\beta_j)\prod_{j=1}^{q}\left(z-\frac{1}{\bar{\beta}_j}\right)$$

$$=c_1\prod_{j=1}^{q}(z-\beta_j)\prod_{j=1}^{q}(z^{-1}-\bar{\beta}_j)$$

这时

$$f(\lambda)=A(e^{i\lambda})$$

$$=c_1\prod_{j=1}^{q}(e^{i\lambda}-\beta_j)\prod_{j=1}^{q}(e^{-i\lambda}-\bar{\beta}_j)$$

$$=c_1\left|\prod_{j=1}^{q}(e^{i\lambda}-\beta_j)\right|^2$$

若记

$$\sqrt{2\pi\frac{c_1}{\sigma^2}}\prod_{j=1}^{q}(e^{i\lambda}-\beta_j)$$

$$=\sum_{k=0}^{q}\bar{b}_k e^{i(q-k)\lambda}=g(e^{i\lambda})$$

则

$$g(z)=\sum_{k=0}^{q}\bar{b}_k z^{q-k}$$

的根的模都不超过 1，而

$$f(\lambda)=\frac{\sigma^2}{2\pi}|g(e^{i\lambda})|^2$$

$$=\frac{\sigma^2}{2\pi}\left|\sum_{k=0}^{q}\bar{b}_k e^{i(q-k)\lambda}\right|^2$$

$$=\frac{\sigma^2}{2\pi}\left|\sum_{k=0}^{q}\bar{b}_k e^{-ik\lambda}\right|^2 \tag{20}$$

即(c)成立.

(c)→(a),若(c)成立,则由上面的证明可知(20)成立,我们只对 $g(\mu)$ 的根的模都小于1时来推出(a). 若 X_T 有如下谱表示,

$$X(n)=\int_{-\pi}^{\pi}e^{in\lambda}d\zeta_X(\lambda)$$

则因为 $g(z)$ 的根在圆内,所以 $g(e^{i\lambda})\neq 0$, 令

$$\tilde{\xi}(n)=\int_{-\pi}^{\pi}\frac{e^{in\lambda}}{g(e^{i\lambda})}d\zeta_X(\lambda)$$

则不难验证 $\tilde{\xi}_T=\{\tilde{\xi}(n), n=0,\pm 1,\cdots\}$ 为平稳过程,且其谱函数为

$$F_{\tilde{\xi}}(\lambda)=\int_{-\pi}^{\lambda}\frac{1}{|g(e^{i\lambda})|^2}f(\lambda)d\lambda$$

$$=\frac{\sigma^2}{2\pi}(\lambda+\pi)$$

所以 $\tilde{\xi}_T$ 是一个白噪声,若 $\tilde{\xi}_T$ 有谱分解

$$\tilde{\xi}(n)=\int_{-\pi}^{\pi}e^{in\lambda}d\zeta_{\tilde{\xi}}(\lambda)$$

则容易推出

$$X(n)=\int_{-\pi}^{\pi}e^{in\lambda}|g(e^{i\lambda})e^{-iq\lambda}|d\zeta_{\tilde{\xi}}(\lambda)$$

$$=\int_{-\pi}^{\pi}\sum_{k=0}^{q}\bar{b}_k e^{i(n-k)\lambda}d\zeta_{\tilde{\xi}}(\lambda)$$

$$=\sum_{k=0}^{q}\bar{b}_k\tilde{\xi}(n-k)$$

即(a)成立. 且由引理1,当 $g(z)$ 的根在圆内时,

$$\tilde{\xi}(n)=\sum_{l=0}^{\infty}c_l X(n-l)\in H_n(X)$$

所以(19)成立，当 $g(z)$ 的根在圆周上时，定理的结论也同样成立[①].

在一般的滑动和表示(16)中，$X(n)\in H_n(\xi)$ 是显然的，而在上面的定理中，指出了必存在一种表示，使 $\xi(n)\in H_n(X)$ 也成立，这样就有

$$H_n(X)=H_n(\xi)$$

这一点在下一章考虑预测问题时是很重要的.

上面我们给出了可表为滑动和过程的协方差函数或谱函数的条件，下面我们来讨论平稳过程可表为一般的 ARMA 过程的各种条件.

在实际应用中考虑 ARMA 过程时，尤其是在预测和控制中，经常把满足(1)的 X_T 看为以 ξ_T 作为输入时系统的输出，而且可以认为 $X(n)$ 只依赖于现在和过去的 ξ_T，即

$$X(n)\in H_n(\xi)$$

这时将来的 $\xi(k)$ 与现在的 $X(n)$ 是不相关的，即

$$\xi(n+1)\perp\{X(k),\ k\leqslant n\}$$

或者写为

$$\xi(n)\perp H_{n-1}(X) \tag{21}$$

若记(1)的右端为

$$\eta(n)\triangleq\sum_{j=0}^{q}b_j\xi(n-j)$$

则

$$\sum_{j=0}^{p}a_jX(n-j)=\eta(n) \tag{22}$$

而由(21)可得

$$\eta(n)\perp H_{n-q-1}(X) \tag{23}$$

反之，当 X_T，η_T 满足(22)，(23)时，因为

① 参阅郑绍濂等校译"希尔伯脱空间平稳序列"，上海科技出版社(1962).

$$\eta(n-q-k)=\sum_{j=0}^{p}a_jX(n-q-k-j)$$

这样，利用(23)，对 η_T 的协方差函数 $B_\eta(n)$ 有

$$B_\eta(q+k)=E\{\eta(n)\overline{\eta(n-q-k)}\}$$
$$=0,\qquad k>0$$

即 η_T 满足定理1的条件，所以由定理1它可表为

$$\eta(n)=\sum_{j=0}^{q}\tilde{b}_j\tilde{\xi}(n-j)$$

其中 $\tilde{\xi}_T$ 为白噪声．因此满足(22)，(23)的 X_T 必可表为(1)，且能使其右端的 $\tilde{b}_j$ 满足

$$B(z)=\sum_{j=0}^{q}\tilde{b}_jz^{q-j}$$

的根全在 $|z|\leqslant 1$ 上．下面我们将讨论满足(22)，(23)的ARMA过程．

定理2 零均值平稳时间序列 X_T 满足(22)，(23)的充要条件是其协方差函数 $B(n)$ 满足

$$\sum_{j=0}^{p}a_jB(n-j)=0,\quad n\geqslant q+1 \tag{24}$$

〔证明〕 必要性．由(22)，(23)

$$\sum_{j=0}^{p}a_jX(n-j)\perp H_{n-q-1}(X)$$

所以当 $n-q-1\geqslant 0$ 时

$$E\left\{\left(\sum_{j=0}^{p}a_jX(n-j)\right)\overline{X(0)}\right\}=0$$

即

$$\sum_{j=0}^{p}a_jB(n-j)=0,\quad n\geqslant q+1$$

充分性. 取

$$\eta(n)=\sum_{j=0}^{p}a_jX(n-j)$$

则$\{\eta(n)\}$也是平稳过程,且

$$E\{\eta(n)\overline{X(m)}\}=\sum_{j=0}^{p}a_jE\{X(n-j)\overline{X(m)}\}$$

$$=\sum_{j=0}^{p}a_jB(n-m-j)$$

所以当$n-m\geqslant q+1$时,$E\{\eta(n)\overline{X(m)}\}=0$, 故

$$\eta(n)\perp H_{n-q-1}(X)$$

即(23)成立.

系 零均值平稳时间序列X_T可表为p阶自回归

$$\sum_{j=0}^{p}a_jX(n-j)=\xi(n) \tag{25}$$

且 $$\xi(n)\perp H_{n-1}(X)$$

的充要条件是

$$\sum_{j=0}^{p}a_jB(n-j)=0,\quad n\geqslant 1 \tag{26}$$

方程(24), (26)在时间序列分析中也称尤尔-瓦尔克 (Yule-Walker)方程, 它在 ARMA 模型的参数a_j的估计中是十分有用的.

定理 3 若零均值平稳时间序列X_T存在谱密度$f(\lambda)$, 则存在白噪声$\xi_T=\{\xi(n),n=0,\ \pm1,\ \cdots\}$使$X_T$可表为ARMA过程(1)的充要条件是$f(\lambda)$可表为

$$f(\lambda)=\frac{\sigma^2}{2\pi}\cdot\frac{\left|\sum_{j=0}^{q}b_je^{-ij\lambda}\right|^2}{\left|\sum_{j=0}^{p}a_je^{-ij\lambda}\right|^2} \tag{27}$$

[证明] 必要性. 若 $X(n)$, $\xi(n)$的谱分解为

$$X(n)=\int_{-\pi}^{\pi} e^{in\lambda}d\zeta_X(\lambda)$$

$$\xi(n)=\int_{-\pi}^{\pi} e^{in\lambda}d\zeta_\xi(\lambda)$$

则

$$\sum_{j=0}^{p}a_jX(n-j)$$

$$=\int_{-\pi}^{\pi} e^{in\lambda}\left(\sum_{j=0}^{p}a_je^{-ij\lambda}\right)d\zeta_X(\lambda) \tag{28}$$

$$\sum_{j=0}^{q}b_j\xi(n-j)=\int_{-\pi}^{\pi} e^{in\lambda}\left(\sum_{j=0}^{q}b_je^{-ij\lambda}\right)d\zeta_\xi(\lambda) \tag{29}$$

又由(1), 对任意的 k,

$$E\left\{\sum_{j=0}^{p}a_jX(n+k-j)\overline{\sum_{j=0}^{p}a_jX(n-j)}\right\}$$

$$=E\left\{\sum_{j=0}^{q}b_j\xi(n+k-j)\overline{\sum_{j=0}^{q}b_j\xi(n-j)}\right\}$$

将(28), (29)代入上式就有

$$\int_{-\pi}^{\pi} e^{i\lambda k}\left|\sum_{j=0}^{p}a_je^{-ij\lambda}\right|^2 f(\lambda)d\lambda$$

$$=\int_{-\pi}^{\pi} e^{i\lambda k}\left|\sum_{j=0}^{q}b_je^{-ij\lambda}\right|^2 \frac{\sigma^2 d\lambda}{2\pi}$$

由 k 的任意性, 可得到

$$f(\lambda)=\frac{\sigma^2}{2\pi}\cdot\frac{\left|\sum_{j=0}^{q}b_je^{-ij\lambda}\right|^2}{\left|\sum_{j=0}^{p}a_je^{-ij\lambda}\right|^2}$$

即(27)成立。

充分性．若(27)成立，令

$$\eta(n)=\sum_{j=0}^{p}a_jX(n-j)$$

则 $\{\eta(n)\}$ 为平稳时间序列，且容易求得 $\eta(n)$ 存在如下谱密度 $f_\eta(\lambda)$:

$$f_\eta(\lambda)=\left|\sum_{j=0}^{p}a_je^{-ij\lambda}\right|^2f(\lambda)$$

$$=\frac{\sigma^2}{2\pi}\left|\sum_{j=0}^{q}b_je^{-ij\lambda}\right|^2$$

由定理1，存在白噪声$\{\xi(n)\}$，使$\eta(n)$可表为

$$\eta(n)=\sum_{j=0}^{q}b_j\xi(n-j)$$

所以对$X(n)$成立

$$\sum_{j=0}^{p}a_jX(n-j)=\sum_{j=0}^{q}b_j\xi(n-j)$$

即 X_T 为 ARMA 过程．

系　若零均值平稳时间序列 X_T 存在谱密度 $f(\lambda)$，则 X_T 可表为 p 阶自回归(25)的充要条件是 $f(\lambda)$ 可表为如下形式

$$f(\lambda)=\frac{\sigma^2}{2\pi}\cdot\frac{1}{\left|\sum_{j=0}^{p}a_je^{-ij\lambda}\right|^2} \tag{30}$$

具有形为(27)，(30)的谱密度又称为**有理谱密度**．上面的两个定理告诉我们，平稳时间序列 X_T 可表为满足(21)的 ARMA 过程时其协方差函数的表征性条件——满足尤尔-瓦尔克方程和存在谱密度时谱密度的表征性条件——具有有理谱．下面我们进一步说明满足(21)的 ARMA 过程其自回归部分的参数 a_j 亦不是任意的．

定理 4　若存在零均值平稳时间序列 X_T 可表为满足 (21) 的 ARMA 过程．则

$$A(z) \triangleq \sum_{j=0}^{p} a_j z^{p-j}$$

在 $|z| \leqslant 1$ 至少有 $p-q$ 个零点，且它在 $|z| \geqslant 1$ 上的零点必是

$$B(z) \triangleq \sum_{j=0}^{q} b_j z^{q-j}$$

和

$$B^*(z) \triangleq \sum_{j=0}^{q} \bar{b}_j z^j$$

的零点．

［证明］　若 $F(\lambda)$ 为 X_T 的谱函数，则因为 X_T 可表为(1)，便有

$$\begin{aligned}
&\int_{-\pi}^{\mu} |A(e^{i\lambda})|^2 dF(\lambda) \\
&= \int_{-\pi}^{\mu} \left|\sum_{j=0}^{p} a_j e^{-ij\lambda}\right|^2 dF(\lambda) \\
&= \int_{-\pi}^{\mu} \left|\sum_{j=0}^{q} b_j e^{-ij\lambda}\right|^2 \frac{\sigma^2 d\lambda}{2\pi} \\
&= \frac{\sigma^2}{2\pi} \int_{-\pi}^{\mu} |B(e^{i\lambda})|^2 d\lambda
\end{aligned}$$

若记 $F_1(\lambda)$ 为 $F(\lambda)$ 的绝对连续部分，且 $f(\lambda) = F_1'(\lambda)$，则

$$\int_{-\pi}^{\mu} |A(e^{i\lambda})|^2 f(\lambda) d\lambda = \int_{-\pi}^{\mu} |B(e^{i\lambda})|^2 d\lambda \cdot \frac{\sigma^2}{2\pi}$$

$$\int_{-\pi}^{\pi} |A(e^{i\lambda})|^2 d(F - F_1) = 0 \tag{31}$$

由 μ 的任意性，必有

$$|A(e^{i\lambda})|^2 f(\lambda) = \frac{\sigma^2}{2\pi} |B(e^{i\lambda})|^2 = \frac{\sigma^2}{2\pi} e^{-qi\lambda} B(e^{i\lambda}) B^*(e^{i\lambda})$$

由于 $f(\lambda)$ 的可积性，所以 $A(z)$ 在 $|z|=1$ 上的零点必是 $B(z)$ 或 $B^*(z)$ 的零点．为叙述简单起见，假定 $A(z)$ 在 $|z|=1$ 上无零点（否则对证明稍加修改即可）且可认为 $B(z)$ 的零点全在 $|z|\leqslant 1$ 上（否则对 $B(z)$，$B^*(z)$ 的零点配置稍加修改即可），则 $B^*(z)$ 的根全在 $|z|\geqslant 1$ 上，且

$$f(\lambda)=\frac{\sigma^2}{2\pi}\left|\frac{B(e^{i\lambda})}{A(e^{i\lambda})}\right|^2$$

$$=\frac{\sigma^2}{2\pi}\frac{B(e^{i\lambda})B^*(e^{i\lambda})}{A(e^{i\lambda})A^*(e^{i\lambda})}e^{i(p-q)\lambda}$$

其中

$$A^*(z)\triangleq\sum_{j=1}^{p}\bar{a}_j z^j$$

又因为 X_T 满足(22)，(23)，由定理 2，其协方差函数满足下列条件

$$\sum_{j=0}^{p}a_j B(n-j)=0,\qquad n\geqslant q+1$$

利用(31)，当 $n\geqslant q+1$ 时可得

$$0=\int_{-\pi}^{\pi}e^{in\lambda}\left(\sum_{j=0}^{p}a_j e^{-ij\lambda}\right)dF(\lambda)$$

$$=\int_{-\pi}^{\pi}e^{i(n-p)\lambda}A(e^{i\lambda})f(\lambda)d\lambda$$

$$+\int_{-\pi}^{\pi}e^{i(n-p)\lambda}A(e^{i\lambda})d(F(\lambda)-F_1(\lambda))$$

$$=\int_{-\pi}^{\pi}e^{i(n-q)\lambda}A(e^{i\lambda})\frac{B(e^{i\lambda})B^*(e^{i\lambda})}{A(e^{i\lambda})A^*(e^{i\lambda})}\cdot\frac{\sigma^2}{2\pi}d\lambda$$

$$=\sigma^2\oint_{|z|=1}z^{n-q-1}\frac{B(z)B^*(z)}{A^*(z)}dz$$

所以由复变函数的知识可以知道，$\dfrac{B(z)B^*(z)}{A^*(z)}$ 在圆内解析，又因为 $B^*(z)$ 的根全在 $|z|\geqslant 1$ 上，所以 $A^*(z)$ 在圆内的零点必是 $B(z)$ 在

圆内的零点，即至多只能有 q 个. 若 z_0 是 $A(z)=0$ 的根，则 $\bar{z}_0^{-1}$ 必是 $A^*(z)=0$ 根，其逆亦真. 所以 $A(z)$ 在圆外至多只能有 q 个零点，而且这些零点必是 $B^*(z)$ 的零点，又由于 $A(z)$ 是 p 次多项式，所以 $A(z)$ 在圆内还至少有 $p-q$ 个零点. 定理证毕.

系　零均值平稳时间序列 X_T 可表为 p 阶自回归(25)，则它满足 $\xi(n)\perp H_{n-1}(X)$ 的充要条件是 $A(z)\triangleq\sum_{j=0}^{p}a_j z^{p-j}$ 的零点全在 $|z|\leqslant 1$ 上. 特别当 $E|\xi(n)|^2\neq 0$ 时，则 $A(z)$ 在零点全在 $|z|<1$ 上.

从上面的定理中可以知道，假若我们是关心的是满足(21)的 ARMA 过程，那么可以认为 $A(z)$ 的零点在 $|z|\leqslant 1$ 上，因为谱密度中对应于 $A(z)$ 在圆外的零点的那一部分是可以与分子 $B(z)B^*(z)$ 相约的，这也意味着 X_T 是一个阶数更低的 ARMA 过程，所以讨论 ARMA 过程时，不论是由(1)给出或由满足(24)的协方差函数或有理谱给出，都可认为是 $A(z)$ 的根的在 $|z|\leqslant 1$ 上，又因为从谱或协方差本身是无法对 $B(z)$ 零点位置加以限制的，但由于在解决预测问题时运用引理 1 的需要，今后我们也常假定 $B(z)$ 的零点在 $|z|\leqslant 1$ 上，这一假定并没对 X_T 的协方差 $B(n)$ 和谱密度 $f(\lambda)$ 有任意附加的限制. 有时，为了避免 $A(z)$，$B(z)$ 零点在 $|z|=1$ 上这一退化的情况，在考虑 ARMA 过程时更要求 $A(z)$，$B(z)$ 的零点全在圆内，这时便有如下结果.

定理 5　若平稳时间序列 X_T 是可表为(1)的 ARMA 过程，则(i)当 $\{a_j\}$ 满足下列条件时，

$$A(z)\triangleq\sum_{j=0}^{p}a_j z^{p-j} \text{ 的零点都在 } |z|<1 \text{ 内} \tag{32}$$

便有

$$X(n)=\sum_{j=0}^{\infty}c_j\xi(n-j)\in H_n(\xi) \tag{33}$$

其中$\sum_{j=0}^{\infty}|c_j|<\infty$;

(ii) 当$\{b_j\}$满足下列条件时,

$$B(z)\triangleq\sum_{j=0}^{p}b_jz^{p-j} \text{ 在圆外 } |z|\geqslant 1 \text{ 无零点} \tag{34}$$

便有

$$\xi(n)=\sum_{j=0}^{\infty}d_jX(n-j)\in H_n(X)$$

其中$\sum_{j=0}^{\infty}|d_j|<\infty$;

(iii) 当(32),(34)同时成立时,$H_n(\xi)=H_n(X)$.

[证明]

(i) 若记

$$\eta(n)\triangleq\sum_{j=0}^{p}a_jX(n-j)$$

$$=\sum_{j=0}^{q}b_j\xi(n-j) \tag{35}$$

则当(32)成立时,由引理1, $X(n-j)=\sum_{l=j}^{\infty}\tilde{c}_l\eta(n-l)$, 且$\sum_{l=0}^{\infty}|\tilde{c}_l|<\infty$. 再将 $\eta(n)=\sum_{j=0}^{q}b_j\xi(n-j)$ 代入上式$\eta(n)$即得(33), 同样可证(ii).

(iii) 因为由(33)

$$X(k)\in H_n(\xi), \qquad k\leqslant n$$

所以$H_n(X)\subset H_n(\xi)$. 又因(34)成立,$H_n(\xi)\subset H_n(X)$,所以

$$H_n(\xi)=H_n(X).$$

最后,我们来说明 ARMA 过程的另一个重要性质.

现在我们引入p维平稳时间序列$Y_T=\{Y(n), n=0,\pm1,\cdots\}$,

$$Y(n)=\begin{bmatrix}Y_1(n)\\ \vdots\\ Y_p(n)\end{bmatrix}$$为 p 维随机变量，满足下列条件：

$$Y(n)=FY(n-1)+\eta(n) \tag{36}$$

$\{\eta(n)\}$ 为 p 维白噪声序列，即

$$E\eta(n)=0$$

$$E\{\eta(n)\eta(m)'\}=G\delta_{mn} \tag{37}$$

其中 $\eta(m)'$ 表 $\eta(n)$ 的共轭转置，则 Y_T 为 p 维的一阶自回归过程。

定理 6 平稳时间序列 X_T 可表示为满足 (32) 的 ARMA 过程(1)，且 $q\leqslant p-1$ 的充要条件是 $\{X(n)\}$ 是满足(36)——(37)的 p 维平稳时间序列的一个分量，且 F 的特征根的模全小于 1。

［证明］ 充分性．在证明充分性时，我们要利用线性代数中下列事实（凯莱-哈密顿（Caylay-Hamilton）定理）对 $p\times p$ 矩阵 F，若

$$\det(F-\mu I)=\sum_{j=0}^{p}a_j\mu^{p-j} \tag{38}$$

则

$$\sum_{j=0}^{p}a_jF^{p-j}=0 \tag{39}$$

现由(36)

$$Y(n-p+1)=FY(n-p)+\eta(n-p+1)$$

$$Y(n-p+2)=F^2Y(n-p)+\eta(n-p+2)+F\eta(n-p+1)$$

……………………………………………………

$$Y(n-j)=F^{p-j}Y(n-p)+\sum_{i=j}^{p-1}F^{i-j}\eta(n-i)$$

……………………………………………………

因此，利用(39)可得

$$\sum_{j=0}^{p} a_j Y(n-j)$$

$$=\left(\sum_{j=0}^{p} a_j F^{p-j}\right) Y(n-p) + \sum_{j=0}^{p-1} a_j \left(\sum_{i=j}^{p-1} F^{i-j} \eta(n-i)\right)$$

$$=\sum_{i=0}^{p-1}\left(\sum_{j=0}^{i} a_j F^{i-j}\right)\eta(n-i) = \sum_{i=0}^{p-1} B_i \eta(n-i)$$

其中$B_i \triangleq \sum_{j=0}^{i} a_j F^{i-j}$，若$X(n)$可表为$Y(n)$的第$l$个分量，那么以$B_i^{(l)}$记$B_i$的$l$行时，就可得到

$$\sum_{j=0}^{p} a_j X(n-j) = \sum_{i=0}^{p-1} B_i^{(l)} \eta(n-i)$$

若记 $\zeta(n) \triangleq \sum_{i=0}^{p-1} B_i^{(l)} \eta(n-i)$，则不难验证$\{\zeta(n)\}$为一维平稳时间序列，且利用 $\zeta(n) = \sum_{i=0}^{p-1} B_i^{(l)} \eta(n-i) \perp H_{n-p}(\eta)$ 容易证明

$$E\{\zeta(n)\overline{\zeta(m)}\} = 0, \quad |n-m| \geqslant p$$

因此由定理1，存在一维白噪声$\{\xi(n)\}$，使$\zeta(n)$可表为

$$\zeta(n) = \sum_{j=0}^{p-1} b_j \xi(n-j)$$

即 X_T 为ARMA 过程且可表为

$$\sum_{j=1}^{p} a_i X(n-j) = \sum_{j=0}^{p-1} b_j \xi(n-j)$$

又因为F的特征根就是(38)中$\sum_{j=0}^{p} a_j \mu^{p-j} \triangleq A(\mu)$的零点，所以(32)成立。

必要性. 若 X_T 满足(32)，则由定理 5 及(33)式，

$$\xi(n+k) \perp X(n) = \sum_{j=0}^{\infty} c_j \xi(n-j), \quad k \geqslant 1 \tag{40}$$

今以 P_G 表二阶矩随机变量希尔伯脱空间中到子空间G的投影算

子，则(40)表示

$$P_{H_n(X)}\xi(n+k)=0,\quad k>0 \tag{41}$$

令

$$Y(n)\triangleq\begin{bmatrix}X(n)\\ P_{H_n(X)}X(n+1)\\ P_{H_n(X)}X(n+2)\\ \vdots\\ P_{H_n(X)}X(n+p-1)\end{bmatrix}$$

则不难验证$\{Y(n), n=0, \pm1, \cdots\}$为$p$维的平稳序列．若以$H_n(Y)$表$\{Y(k),\ k\leqslant n\}$诸分量的任意线性组合及其均方极限张成的子空间，则$H_n(X)=H_n(Y)$．令$P_n=P_{H_n(X)}=P_{H_n(Y)}$，因为$H_n(X)\subset H_{n+1}(X)$，所以$P_nP_{n+1}=P_n$，且

$$Y(n)=\begin{bmatrix}X(n)\\ P_nX(n+1)\\ \vdots\\ P_nX(n-p-1)\end{bmatrix}$$

$$Y(n+1)=\begin{bmatrix}X(n+1)\\ P_{n+1}X(n+2)\\ \vdots\\ P_{n+1}X(n+p)\end{bmatrix}$$

若以$Y_l(n)$表$Y(n)$的第l个分量，我们来考察$P_nY(n+1)$，

$$\left.\begin{aligned}&P_nY_1(n+1)=P_nX(n+1)=Y_2(n)\\ &P_nY_2(n+1)=P_nP_{n+1}X(n+2)=P_nX(n+2)=Y_3(n)\\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\\ &P_nY_p(n+1)=P_nP_{n+1}X(n+p)=P_nX(n+p)\\ &=P_n\Big(-\sum_{j=1}^{p}a_jX(n+p-j)+\sum_{j=0}^{q}b_j\xi(n+p-j)\Big)\\ &=-\sum_{j=1}^{p}a_jP_nX(n+p-j)=-\sum_{j=1}^{p}a_jY_{p-j+1}(n)\end{aligned}\right\}\tag{42}$$

在上面的推导中，我们依据 $q\leqslant p-1$ 及(40)而利用了

$$P_{\Lambda}\left(\sum_{j=0}^{q}b_j\xi(n+p-j)\right)=0$$

由(42)

$$P_{\Lambda}Y(n+1)=FY(n)$$

其中 F 为

$$F=\begin{bmatrix}0 & 1 & 0 & \cdots & 0\\ 0 & 0 & 1 & \cdots & 0\\ \cdots & \cdots & \cdots & \cdots & \cdots\\ 0 & 0 & 0 & \cdots & 1\\ -a_p & -a_{p-1} & -a_{p-2} & \cdots & -a_1\end{bmatrix}$$

若记

$$Y(n+1)=FY(n)+\eta(n+1) \tag{43}$$

那么不难验证$\{\eta(n)\}$是 p 维平稳序列，而且依据投影定理

$$\eta(n+1)\perp H_n(Y) \tag{44}$$

又由(43)$\eta(n)\in H_n(Y)$，这样就有

$$\eta(n+1)\perp\{\eta(n),\eta(n-1),\cdots\}$$

所以 $\{\eta(n)\}$ 还是一个 p 维白噪声，而 F 的特征方程就是 $A(\mu)=\sum_{j=0}^{p}a_j\lambda^{n-j}(a_0=1)$，由(32)保证的 F 特征根的模全小于 1，定理证毕.

在定理 6 中对 X_T 表为(1)时要求 $q\leqslant p-1$ 不是一个本质性的限制，因为当 $q\geqslant p$ 时，我们总可以令 $a_{p+1}=a_{p+2}=\cdots=a_{q+1}=0$ 把 X_T 表为如下，具有 $p\geqslant q+1$ 的形式

$$\sum_{j=0}^{q+1}a_jX(n-j)=\sum_{j=0}^{q}b_j\xi(n-j)$$

所以只要 $A(z)$的零点在圆内，ARMA 过程就是多维一阶自回归过程的分量.

从上面的定理证明中可以看出，Y_T 为正态时，X_T 也必定是正态的，而 X_T 为正态时，Y_T 也可取为是正态的，而当 Y_T 是正态过程时，又因为它满足(44)，可以像一维情形一样证明它还是一个马尔可夫过程，所以一个正态 ARMA 过程当 $A(z)$ 的零点在圆内时它还必定是一个多维马尔可夫过程的分量.

至此我们较为详细地讨论了 ARMA 过程的性质，当过程的参数是连续变化时，也可类似地用微分方程式代替差分方程来规定 ARMA 过程，但要涉及到连续参数白噪声的导数，相应的讨论就要复杂一些①。

习　题

1. 若$\{\eta(t), -\infty<t<\infty\}$为零均值的正交增量过程，$E|\eta(t_2)-\eta(t_1)|^2=t_2-t_1$. 试证

$$X(t)=\eta(t)-\eta(t-1)$$

为平稳过程，并求出它的协方差函数和谱函数.

2. 若$\{X(t)\}$, $\{Y(t)\}$是平稳相关的平稳过程，且其均值函数，协方差函数 $B_X(\tau)$, $B_Y(\tau)$和互协方差函数 $B_{XY}(\tau)$满足:

$$E\{X(t)\}=0, \qquad E\{Y(t)\}=0$$

$$B_X(\tau)=B_Y(\tau), \qquad B_{XY}(\tau)=-B_{XY}(-\tau)$$

试证 $W(t)=X(t)\cos\omega_0 t+Y(t)\sin\omega_0 t$ 也是平稳过程. 又若 X, Y 的谱密度存在，则 W 也存在谱密度，试用 X, Y 的谱密度与交互谱密度表出 $W(t)$ 的谱密度.

3. 若 $X(t)$ 为平稳正态过程，$E\{X(t)\}=0$, $B(\tau)$ 为其协方差函数. 又

$$Y(t)=\operatorname{sgn}[X(t)]=\begin{cases}+1, & \text{当 } X(t)\geqslant 0\\ -1, & \text{当 } X(t)<0\end{cases}$$

试证 $Y(t)$ 是平稳过程，且其标准相关函数为

$$R_Y(\tau)=\frac{2}{\pi}\arcsin\left(\frac{B(\tau)}{B(0)}\right)$$

---

① 参见郑绍濂等编校"希尔伯脱空间中的平稳序列"，上海科技出版社 (1962年)，第三篇文章.

4. 若 $\{X(t), -\infty<t<\infty\}$ 为均方连续平稳过程，具有谱密度 $f(\lambda)$. 试证对每个 $\Delta>0$, $X^{(\Delta)}=\{X(n\Delta), n=0, \pm1, \cdots\}$ 为平稳序列，并用 $f(\lambda)$ 表示出 $X^{(\Delta)}$ 的谱密度.

5. 若下列 $f(\lambda)$ 为平稳序列 $\{X(n), n=0, \pm1, \cdots\}$ 的谱密度，验证相应的 $B(n)$ 为 $\{X(n), n=0, \pm1, \cdots\}$的协方差序列:

(i) $f(\lambda)=\dfrac{1}{4\pi}\left[\dfrac{1-|a|^2}{|1-ae^{i\lambda}|^2}+\dfrac{1-|a|^2}{|1-\bar{a}e^{i\lambda}|^2}\right], \quad a=re^{i\varphi},$

$$B(n)=r^{|n|}\cos n\varphi$$

(ii) $f(\lambda)=\dfrac{|e^{i\lambda}-b|^2}{2\pi|e^{i\lambda}-a_1|^2|e^{i\lambda}-a_2|^2}, \quad |b|\leqslant1, \quad |a_1|<1, \quad |a_2|<1,$

$$B(n)=\begin{cases}\dfrac{1}{a_1-a_2}\left[\dfrac{a_1^{|n|}(a_1-b)(1-a_1\bar{b})}{(1-|a_1|^2)(1-a_1\bar{a}_2)}-\dfrac{a_2^{|n|}(a_2-b)(1-a_2\bar{b})}{(1-|a_2|^2)(1-\bar{a}_1a_2)}\right], & a_1\neq a_2\\ \dfrac{a^{|n|}(1+|a|^2+|b|^2-4\mathrm{Re}(\bar{a}b)+|a|^2|b|^2)}{(1-|a|^2)^3} & \\ +\dfrac{|n|(1-a\bar{b})(a-b)a^{|n|-1}}{(1-|a|^2)^2}, & a_1=a_2=a\end{cases}$$

并写出 a_2 或(和)b 为零时，对应的 $f(\lambda)$ 和 $B(n)$.

6. 若下列 $f(\lambda)$ 为平稳过程 $\{X(t), -\infty<t<\infty\}$ 的谱密度，验证相应的 $B(t)$为$\{X(t), -\infty<t<\infty\}$的协方差函数

(i) $f(\lambda)=\dfrac{\alpha[\lambda^2+(\alpha^2+\beta^2)]}{\pi[\lambda^4+2(\alpha^2-\beta^2)\lambda^2+(\alpha^2+\beta^2)^2]}, \ \alpha, \beta>0,$

$B(t)=e^{-\alpha|t|}\cos\beta t$

(ii) $f(\lambda)=\dfrac{|i\lambda-b|^2}{2\pi|i\lambda-a_1|^2|i\lambda-a_2|^2}, \mathrm{Re}a_1<0, \mathrm{Re}a_2<0,$

$$B(t)=\begin{cases}\dfrac{1}{2(a_2-a_1)}\left[\dfrac{(a_1-b)(a_1+\bar{b})e^{a_1|t|}}{\mathrm{Re}a_1(a_1+\bar{a}_2)}-\dfrac{(a_2-b)(a_2+\bar{b})e^{a_2|t|}}{\mathrm{Re}a_2(a_2+\bar{a}_1)}\right], & a_1\neq a_2\\ e^{a|t|}\left[\dfrac{|a|^2+|b|^2-2\mathrm{Im}a\cdot\mathrm{Im}b}{4(\mathrm{Re}a)^3}-\dfrac{(a-b)(a+\bar{b})}{4(\mathrm{Re}a)^2}|t|\right] & \end{cases}$$

并写出 b 或 a_2, b 为零时，对应的 $f(\lambda)$和 $B(t)$.

7. 试证若平稳过程 X_T 的谱密度 $f(\lambda)$可表为

$$f(\lambda)=\frac{c_1}{2\pi}+\frac{c_2}{2\pi|e^{i\lambda}-a|^2}, \qquad c_1, c_2>0$$

的充要条件是 $f(\lambda)$可表为

$$f(\lambda)=\frac{|e^{i\lambda}-b|^2}{2\pi|e^{i\lambda}-a|^2},\ |a|<1,\ |b|<1,\ |b|<|a|$$

这时也必存在白噪声序列$\{\xi(n)\}$和马尔可夫序列$\{\eta(n)\}$，使$X(n)$和$\xi(n)+\eta(n)$有相同的协方差函数。

8. 若平稳过程$X(t)$的谱函数$F(\lambda)$对某个a满足

$$\int_{-\infty}^{\infty}e^{|\lambda a|}dF(\lambda)<\infty$$

则$X(t)$任意次均方可微。

9. 若平稳过程$\{X(t),\ -\infty<t<\infty\}$的谱函数不是一个只含有限个跳跃点的纯跳跃函数，则对任意的n及$t_1,t_2,\cdots,t_n$，由协方差函数$B(t)$构成的矩阵

$$(B(t_i-t_j))_{1\leqslant i,\ j\leqslant n}$$

是正定的。

10. 若实平稳过程为狭带信号，即它存在谱密度且满足

$$f(\lambda)=0,\qquad |\lambda|\geqslant\lambda_0$$

试证其协方差函数当$|\tau|\leqslant\frac{\pi}{2\lambda_0}$时满足

$$B(\tau)\geqslant B(0)\cos(\lambda_0\tau)$$

11. 若$X(t)$为平稳过程，且存在均方导数$X'(t)$，试验证下列$Y(t)$也是平稳过程，并用X_T的谱函数表出Y_T的谱函数。

(i) $Y(t)=\int_{-\infty}^{t}e^{-\beta(t-s)}X'(s)ds$;

(ii) $Y(t)=\int_{-\infty}^{t}e^{-a(t-s)}\frac{\sin\omega(t-s)}{\omega}X'(s)ds$.

12. 若$\{\eta(t),-\infty<t<\infty\}$是零均值的正交增量过程，$E|\eta(t_2)-\eta(t_1)|^2=t_2-t_1$，试证

$$\xi(t)=\int_{-\infty}^{\infty}\frac{e^{it\tau}-1}{i\tau\sqrt{2\pi}}d\eta(\tau),\ -\infty<t<\infty$$

也是零均值正交增量过程，且

$$E|\xi(t_2)-\xi(t_1)|^2=|t_2-t_1|$$

$$\eta(t_2)-\eta(t_1)=\int_{-\infty}^{\infty}\frac{e^{-it_2\lambda}-e^{-it_1\lambda}}{i\lambda\sqrt{2\pi}}d\xi(\lambda)$$

13. 若$Y(t)$, $X(t)$为均方连续的平稳过程，且满足

$$Y''(t)+\beta Y'(t)+\omega_0^2Y(t)=X(t)$$

试用 $X(t)$ 的谱函数来表示 $Y(t)$ 的谱函数。

14. 设 $\{X(n)\}$ 为平稳序列，令

$$e(n)=X(n)+\sum_{j=1}^{p}a_jX(n-j)$$

若

$$E\{\varepsilon(n)\overline{X(n-j)}\}=0,\ 1\leqslant j\leqslant p$$

试证 $\varepsilon(n)\perp H_{n-1}(X)$ 的充要条件是 $\{\varepsilon(n)\}$ 为白噪声序列。

15. 若平稳过程 X_T 的谱密度为 $f(\lambda)$，试证

$$E\left|\int_a^b g(t)X(t)dt\right|^2\leqslant 2\pi\sup_\lambda f(\lambda)\int_a^b|g(t)|^2dt$$

16. 若平稳过程 $X_T=\{X(n),n=0,\pm1,\cdots\}$ 具有谱密度 $f(\lambda)$，$\zeta(\lambda)$ 为 X_T 的随机谱函数。又 $H_0(X)$ 表 $\{X(n),n\leqslant0\}$ 张成的 H 的子空间，当 $f(\lambda)\leqslant M$ 时，

$$H(X)\supset\left\{\int_{-\pi}^{\pi}\varphi(\lambda)d\zeta(\lambda);\ \varphi(\lambda)=\sum_{n=-\infty}^{\infty}c_ne^{in\lambda},\ \sum_{n=-\infty}^{\infty}|c_n|^2<\infty\right\}$$

$$H_0(X)\supset\left\{\int_{-\pi}^{\pi}\varphi(\lambda)d\zeta(\lambda);\ \varphi(\lambda)=\sum_{n=-\infty}^{0}c_ne^{in\lambda},\ \sum_{n=-\infty}^{0}|c_n|^2<\infty\right\}$$

而当 $f(\lambda)\geqslant m>0$ 时，成立

$$H(X)\subset\left\{\int_{-\pi}^{\pi}\varphi(\lambda)d\zeta(\lambda);\ \varphi(\lambda)=\sum_{n=-\infty}^{\infty}c_ne^{in\lambda},\ \sum_{n=-\infty}^{\infty}|c_n|^2<\infty\right\}$$

$$H_0(X)\subset\left\{\int_{-\pi}^{\pi}\varphi(\lambda)d\zeta(\lambda);\ \varphi(\lambda)=\sum_{n=-\infty}^{0}c_ne^{in\lambda},\ \sum_{n=-\infty}^{0}|c_n|^2<\infty\right\}$$

第五章　时间序列分析

在这一章里，我们主要对参数集为离散的二阶矩过程 $X_T=\{X(n),\ n=0,\pm1,\cdots\}$ 进行讨论. 因为在利用数字计算机进行处理时，往往需要对参数为连续变化的随机过程进行采样，把它变换为一个参数为离散的随机序列. 所以对离散参数的随机序列进行讨论是十分重要的. 离散参数随机序列使用的许多统计方法也就是计算机上分析连续参数随机过程的方法，而且在处理连续参数随机过程中在随机过程积分等方面遇到的一些数学上的困难，在离散参数随机序列的处理中都可避免. 所以我们在这一章中主要讨论参数集为整数的情形，即考虑的是随机序列.

由于科学技术的需要和计算机技术的发展，本章介绍的时间序列统计分析方法在实用中有很多用处，而且也不局限用于很“正规”的随机过程的统计之中，例如语音声波分析，地震波分析，振动分析等，在那里所处理的时间序列 $\{x(n),\ n=0,1,2,\cdots\}$ 未必可看作为我们以前介绍的随机过程，但是本章中将要介绍的一些适用于随机过程的统计分析方法，在这些领域中仍被广泛的使用，而且获得很大的成功. 所以我们采用“时间序列”这一名词，它可以包括随机过程(主要是参数为整数的随机时间序列)，也可以包括其它波形的数据. 不过因为对时间序列所使用的一些行之有效的统计方法，大多还只能从随机过程的统计分析中找到合理的理论基础，所以我们虽然采用含义广泛的名词“时间序列”，但在这里只是对随机时间序列来讨论它的预测，滤波和其它统计分析问题.

§1. 时间序列的预测和滤波

一、问题的提出

在研究随机现象时，由于不同随机变量之间的统计联系，往往希望利用这种联系的统计规律性，由一些随机变量的观察值去估计或预报另一些随机变量的取值．气象预报中的统计预报就是一例．在上一篇数理统计中的回归分析中就曾介绍过这类问题的处理方法．对于随机过程中也有类似的情况．因为对随机过程来说，它是一族随时间而变化的随机变量，而且一般来说，对应不同时间的随机变量之间是存在着统计联系的，因此我们可以利用对应不同时间的随机变量之间的统计联系，由“过去”的观察值来估计或预报“现在”或“未来”的取值．这类问题在实用中是经常会遇到的．例如在火炮控制或制导系统中，为了有效地控制武器系统向目标攻击，就需要在每个时刻根据目标运动的历史知识来预报出下一个时刻的状态(包括位置、速度等)，以便指挥武器系统准确地指向这一目标．虽然目标的运动是满足一定运动方程的，但不可避免的有随机因素存在，更重要的由于对目标观察存在噪声，所以我们所能获得的目标状态就应该用一个随机过程来描述．而能否准确地预报这一随机过程的取值，就直接影响到控制系统的命中率．

又如在工业生产过程的控制问题中，为了使我们关心的质量或数量指标达到预定的目标，利用系统状态的观察来预报下一时刻的状态，并依据这一点来确定控制方案，这种做法也是常用的．由于系统本身的随机干扰和观测噪声的存在，在这类问题中所处理的也是一个随机过程．

从上面的实际问题中，我们可以把问题归结如下：

若 $X_T=\{X(t), t\in T\}$ 是我们关心的随机过程，一般地它表示一个系统的状态，$X(t)$本身可以是一维的，也可以是多维的，但由于实际观测条件的限制只能观察到$X(t)$的分量或其组合，所以我们可以获得的观察值是

$$Z(t)=H(t)X(t)+V(t) \tag{1}$$

其中$V(t)$是观测的噪声，当$X(t)$为N维向量时，$H(t)$便是一个$m\times N$ 矩阵，是表示我们可以观测到的$X(t)$的某些分量或其线性组合，(1)又称为**观测方程**．我们的问题是如何利用 $Z(t)$ 在$[t_1, t_2]$的观测值来估计出在时刻 $t=t_0$时 $X(t_0)$ 的取值．而观测区间(t_1, t_2)可以是有限的，也可以是无限的．一般地，

当 $t_0=t_2$时，对$X(t_0)$的估计问题又称为**滤波**；

当 $t_0>t_2$时，又称为**预测或外推**；

当 $t_1<t_0<t_2$时，又称为**内插**．

这三类问题统称为**估计问题**．当然，这里的估计是和数理统计中的参数估计是含义不同的，参数估计是对未知参数的估计，而这里是对未知随机变量的估计．

在上述估计问题中，允许利用的是$\{Z(t), t_1\leqslant t\leqslant t_2\}$．所以我们为了对$X(t_0)$进行估计，就必须作出$Z(t)$，$t_1\leqslant t\leqslant t_2$的某个函数 $f(Z(t), t_1\leqslant t\leqslant t_2)$，用它来估计$X(t_0)$．但是在提出估计问题的许多实际问题中，所处理的随机过程除非是正态过程，否则大多是无法确切了解其全部分布函数的知识的，而要求了解其均值，协方差函数等二阶结构是可能的，所以用$Z(t)$，$t_1\leqslant t\leqslant t_2$的任意函数作为估计量时，其好坏就很难讨论，而只能对用$Z(t)$，$t_1\leqslant t\leqslant t_2$的有限线性组合$\sum_{i=1}^{n}c_iZ(s_i)$ 及其均方极限来作为估计量时，讨论其统计性质．这种估计也称为**线性估计**．在$X(t_0)$的不同线性估计$\hat{X}(t_0)$中，我们希望寻求的是使下列均方偏差为最小的估

计，即要求

$$E|\hat{X}(t_0)-X(t_0)|^2=\min \tag{2}$$

对上述估计我们也借用**最小方差估计**的名称。

综上所述，我们的问题便是求：

$$\hat{X}(t_0)\in H(Z)=\{\sum_{i=1}^{n}c_iZ(s_i)\text{及其均方极限，}$$

$$c_i\text{ 任意数，}t_1<s_i\leqslant t_2\} \tag{3}$$

使

$$E|\hat{X}(t_0)-X(t_0)|^2=\min$$

在求解上述问题时，我们总认为过程 X_T 和 Z_T 的二阶统计特性是已知的，即下列函数是已知的，

$$B_{XX}(s,t)=E\{(X(s)-EX(s))(\overline{X(t)-EX(t)})\} \tag{4}$$

$$B_{XZ}(s,t)=E\{(X(s)-EX(s))(\overline{Z(t)-EZ(t)})\} \tag{5}$$

$$B_{ZZ}(s,t)=E\{(Z(s)-EZ(s))(\overline{Z(t)-EZ(t)})\} \tag{6}$$

当然，在实际应用上使用均方偏差最小的线性估计来解决估计问题也是方便而合用的。进一步的分析还表明，当 $X(t_0)$ 和 Z_T 是联合正态过程时，在 $Z(t)$，$t_1\leqslant t\leqslant t_2$ 的一切（可测）函数中考虑 $X(t_0)$ 的估计时，必有线性估计达到最小均方偏差①。所以对正态过程来说，虽然从二阶统计特征可以了解它的全部分布的知识，有可能考虑 $X(t_0)$ 更为一般的估计，但若从均方偏差最小而论，线性估计已是足够了。

二、投影定理及其应用

在线性估计问题中，我们涉及的是二阶矩过程 X_T 和 Z_T。允许使用的估计量是 $H(Z)=\{\sum_{i=1}^{n}c_iZ(s_i)$ 及其均方极限$\}$，它是 $H\triangleq\{\xi,\ E|\xi|^2<\infty\}$ 的一个子空间，而估计量为最佳的准则又是

① 参见中国科学院数学研究所编著"离散时间系统滤波的数学方法"，1975.

(2)，若在H中，记

$$(\xi,\eta)=E(\xi\bar{\eta}),\quad \|\xi\|=(E|\xi|^2)^{\frac{1}{2}},\quad \xi,\eta\in H \tag{7}$$

那么(2)就是找$\hat{X}(t_0)\in H(Z)$使

$$\|X(t_0)-\hat{X}(t_0)\|=\min_{\eta\in H(Z)}\|X(t_0)-\eta\| \tag{8}$$

所以用H中的术语来说，线性估计问题(3)就是要在H的子空间$H(Z)$中找一元素$\hat{X}(t_0)$，使$X(t_0)$和$\hat{X}(t_0)$有最短距离，为了解决这一问题，希尔伯脱空间中的投影定理是有用的.

定理1 （投影定理） 若H_1是希尔伯脱空间H的子空间，对任一$\xi\in H$，存在唯一的元素$\xi_1\in H_1$，满足

$$\|\xi-\xi_1\|=\min_{\eta\in H_1}\|\xi-\eta\| \tag{9}$$

且这时ξ_1和$\xi_2\triangleq\xi-\xi_1$是唯一满足下列条件的ξ的正交分解，

$$\xi=\xi_1+\xi_2,\ \xi_1\in H_1,\ \xi_2\perp H_1 \tag{10}$$

我们可从图1看一下它的几何意义，若$\|\xi-\eta\|$表示向量ξ到H_1中向量η的距离，则投影定理表明H_1中的ξ_1与ξ有最短距离，且$\xi_2=\xi-\xi_1$与H_1正交．所以ξ_1就是ξ在H_1上的投影，ξ_2是ξ到H_1上的垂线.

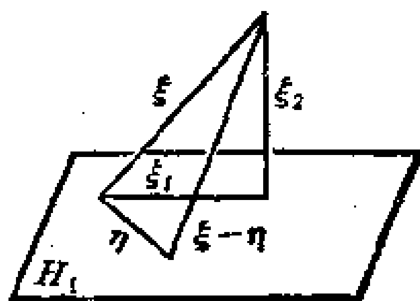


图 1

今后，我们常把H空间中的元素ξ在其子空间H_1上的投影ξ_1记为$P_{H_1}\xi$．若不会引起混淆时也简写为$P\xi$．对于投影运算，我们列举下面两个常用的性质.

性质1 P_{H_1}是线性运算，即对任意复数α,β及H中元素ξ,η：

$$P_{H_1}(\alpha\xi+\beta\eta)=\alpha P_{H_1}\xi+\beta P_{H_1}\eta \tag{11}$$

性质2 若H_2是H_1的子空间，则对任一$\xi\in H$有

$$P_{H_1}P_{H_1}\xi=P_{H_1}\xi=P_{H_1}P_{H_1}\xi \tag{12}$$

利用投影定理立即可推出下列定理.

定理 2 若 $X_T=\{X(t),\ t\in T\}$和 $Z_T=\{Z(t),\ t_1<t\leqslant t_2\}$为二阶矩过程,则(3)中的 Z_T 对 $X(t_0)$的最小方差估计 $\hat{X}(t_0)$为

$$\hat{X}(t_0)=P_{H(Z)}X(t_0) \tag{13}$$

其中$H(Z)$亦由(3)规定. 或者等价地 $\hat{X}(t_0)$ 是满足下列条件的$H(Z)$中的元素

$$E\{\hat{X}(t_0)\overline{Z(t)}\}=E\{X(t_0)\overline{Z(t)}\},\quad t_1<t\leqslant t_2 \tag{14}$$

[证明] 定理的前一半就是投影定理的结论. 至于(14), 我们可以这样来看,当(13)成立时,则由(10)

$$X(t_0)-\hat{X}(t_0)\perp H(Z) \tag{15}$$

所以

$$X(t_0)-\hat{X}(t_0)\perp Z(t),\qquad t_1<t\leqslant t_2 \tag{16}$$

即(14)成立. 反之,当(14)成立时,则

$$(X(t_0)-\hat{X}(t_0),Z(t))=0,\qquad t_1<t\leqslant t_2$$

故当

$$\eta=\sum_{i=1}^{n}c_iZ(s_i),\ t_1<s_i\leqslant t_2,$$

$$(X(t_0)-\hat{X}(t_0),\eta)=0$$

进而对一切 $\eta\in H(Z)$,

$$(X(t_0)-\hat{X}(t_0);\eta)=0$$

同时 $\hat{X}(t_0)\in H(Z)$,所以由正交分解的唯一性,必有

$$\hat{X}(t_0)=P_{H(Z)}X(t)$$

(14) 是最小方差估计所满足的一个方程式, 为了由它求得$\hat{X}(t_0)$还必须根据过程 Z_T 和$X(t_0)$协方差的具体形式, 由(14)推出更明确的表示式, 用来求解 $\hat{X}(t_0)$. 下面我们考虑有限点观测值的简单情形.

［例 1］ 若已知 $Z_T=\{Z(t_i),\ 1\leqslant i\leqslant n\}$，要求 $X(t_0)$ 的最小方差估计．在这个问题中，因为参数集有限，所以

$$H(Z)=\{\sum_{i=1}^{n}c_iZ(t_i),\ c_i\ \text{为任意复数}\} \tag{17}$$

因此 $\hat{X}(t_0)$ 可表为 $\hat{X}(t_0)=\sum_{i=1}^{n}c_iZ(t_i)$．为了求出 c_i 可以利用(14)．在目前的情况下，(14)可写为

$$E\{\sum_{i=1}^{n}c_iZ(t_i)\overline{Z(t_j)}\}=E\{X(t_0)\overline{Z(t_j)}\}$$

即

$$\sum_{i=1}^{n}c_i\Gamma_Z(t_i,t_j)=\Gamma_{XZ}(t_0,t_j),\ 1\leqslant j\leqslant n \tag{18}$$

其中

$$\Gamma_Z(t_i,t_j)=E\{Z(t_i)\overline{Z(t_j)}\}$$
$$\Gamma_{XZ}(t_0,t_j)=E\{X(t_0)\overline{Z(t_j)}\}$$

方程(18)就成为我们求解 c_j 的方程式．特别当 $\{Z(t_i),1\leqslant i\leqslant n\}$ 线性无关时，

$$G\triangleq\det(\Gamma_Z(t_i,t_j))\neq 0 \tag{19}$$

那么由(19)可解出唯一的一组 c_j，$1\leqslant j\leqslant n$，并且不难由此推出

$$\hat{X}(t_0)=-\frac{1}{G}\begin{vmatrix}0 & \Gamma_{XZ}(t_0,t_1) & \cdots & \Gamma_{XZ}(t_0,t_n)\\ Z(t_1) & \Gamma_Z(t_1,t_1) & \cdots & \Gamma_Z(t_1,t_n)\\ \cdots & \cdots & \cdots & \cdots\\ Z(t_n) & \Gamma_Z(t_n,t_1) & \cdots & \Gamma_Z(t_n,t_n)\end{vmatrix} \tag{20}$$

同时可求得估计误差为：

$$E|X(t_0)-\hat{X}(t_0)|^2=E\{(X(t_0)-\hat{X}(t_0))\overline{X(t_0)}\}$$

$$=\frac{1}{G}\begin{vmatrix}E\{X(t_0)\overline{X(t_0)}\} & \Gamma_{XZ}(t_0,t_1) & \cdots & \Gamma_{XZ}(t_0,t_n)\\ \Gamma_{ZX}(t_1,t_0) & \Gamma_Z(t_1,t_1) & \cdots & \Gamma_Z(t_1,t_n)\\ \cdots & \cdots & \cdots & \cdots\\ \Gamma_{ZX}(t_n,t_0) & \Gamma_Z(t_n,t_1) & \cdots & \Gamma_Z(t_n,t_n)\end{vmatrix} \tag{21}$$

所以对有限点观测值的情形，利用(14)是很容易求得估计问题的解的．当进行观测的时间 t 是无限集时，问题就要复杂得多，因为这时 $H(Z)$ 中的元素就不象(17)中规定的那样简单，而(14)的求解更困难．例如已知 $Z_T=\{Z(t), a\leqslant t\leqslant b\}$，这时即使我们假定最佳估计有如下形式

$$\hat{X}(t_0)=\int_a^b h(s)Z(s)dt$$

$h(t)$是一未知的函数．那么由(14)可得求解 $h(t)$的如下积分方程

$$\int_a^b \Gamma_Z(t, s)\overline{h(s)}ds=\Gamma_{ZX}(t, t_0) \tag{22}$$

除非 $\Gamma_Z(t, s)$ 有一些特殊性质可利用(例如它是对应于某个微分子算子的格林函数)或直接进行数值解，否则一般地求解(22)将是十分困难的．下面我们将对较为特殊的时间序列来讨论其预测和滤波问题．

三、无噪声观测平稳序列的预测

现在我们考虑对平稳时间序列 $X_T=\{X(n), n=0, \pm 1, \cdots\}$ 的预测问题，在这里假定获得的观测是无噪声的，即

$$Z(n)=X(n), n=0, \pm 1, \cdots$$

假若已获得的观测值是

$$X(n), n\leqslant 0$$

而要求预测的是 $X(n_0), n_0>0$，标准仍然是均方误差最小．

为简单计，我们假定 $EX(n)=0$．

1．有限点观测值已知的预测

若已知的观察值为

$$X(-n+1), X(-n+2), \cdots, X(-1), X(0)$$

用它来对 $X(n_0)$ 进行预测．若预测值 $\hat{X}(n_0)=\sum_{k=0}^{n-1}c_kX(-k)$，那么在这一情况下方程(18)就可写为

$$\sum_{k=0}^{n-1} c_k B(i-k) = B(n_0+i),\ 0 \leqslant i \leqslant n-1 \tag{23}$$

其中 $B(n)$ 为 X_T 的协方差函数．方程(23)的系数矩阵为

$$\begin{bmatrix} B(0) & B(-1) & \cdots & B(-n+1) \\ B(1) & B(0) & \cdots & B(-n+2) \\ \cdots & \cdots & \cdots & \cdots \\ B(n-1) & B(n-2) & \cdots & B(0) \end{bmatrix}$$

它是一个非负定矩阵，且与主对角线平行的各位置上诸元素都相同，它又称为托布里兹(Toeplitz)矩阵，利用托布里兹矩阵的特性，可使(23)的求解有较快的递推算法①．按(21)也不难写出预测误差为

$$\sigma_{n_0}^2 \triangleq E|X(n_0) - \hat{X}(n_0)|^2$$
$$= \frac{G_{n_0}(n)}{G(n)}$$

其中 $$G(n) = \det(B(i-j))_{i,j=1}^{n}$$

$$G_{n_0}(n) = \det \begin{bmatrix} B(0) & B(n_0)\cdots & B(n_0-n+1) \\ B(-n_0) & B(0)\cdots & B(-n+1) \\ \cdots & \cdots & \cdots \\ B(-n_0+n-1) & B(n-1)\cdots & B(0) \end{bmatrix}$$

所以对有限点已知的预测问题是完全可以解决的．

下面我们来考虑 $X(n)$，$n \leqslant 0$ 都已知的预测问题．考虑这种问题的一个原因是因为它可以作为已知资料足够多的情形的一种近似．

2．自回归模型

若平稳序列 $X_T = \{X(n),\ n=0,\ \pm 1,\ \cdots\}$ 可表为如下自回归形式

---

① 参见 G. E. P. Box, G. M. Jenkins, Time Series Analysis, 1972.

$$\sum_{j=0}^{p} a_j X(n-j) = \xi(n), \quad a_0 = 1 \tag{24}$$

其中 $\xi(n)$为白噪声，并且假定 a_j 满足

$$\sum_{j=0}^{p} a_j \lambda^{p-j} = 0 \tag{25}$$

的根全在单位圆$|\lambda|<1$之内. 这时由定理 4.4.5,

$$H_n(X) = H_n(\xi)$$

$$\xi(n+k) \perp H_n(X), \qquad k>0$$

又由引理 4.4.1, 取 $k=n-1$, 并利用 $X(n)$的平稳性可得

$$X(n) = \sum_{l=0}^{n-1} c_l \xi(n-l) + \sum_{j=1}^{p} b_j X(1-j)$$

所以利用$\{X(k), k\leqslant 0\}$对 $X(n_0)$的最佳线性预测 $\hat{X}(n_0)$为

$$\begin{aligned}\hat{X}(n_0) &= P_{H_0(X)} X(n_0) \\ &= \sum_{l=0}^{n_0-1} c_l P_{H_0(X)} \xi(n_0-l) \\ &\quad + \sum_{j=1}^{p} b_j P_{H_0(X)} X(1-j) \\ &= \sum_{j=1}^{p} b_j X(1-j)\end{aligned} \tag{26}$$

在此我们利用了投影算子的线性性, $\xi(n_0-l)\perp H_0(X)$和$X(1-j)\in H_0(X)$. 特别当 $n_0=1$ 时, 有

$$\begin{aligned}\hat{X}(1) &= P_{H_0(X)} X(1) \\ &= P_{H_0(X)}\left(-\sum_{j=1}^{p} a_j X(1-j) + \xi(1)\right) \\ &= -\sum_{j=1}^{p} a_j X(1-j)\end{aligned} \tag{27}$$

(26)告诉我们对满足(25)的自回归过程(24), 关于$X(n_0)$的预测只依赖"最近的" p 个观察值$X(0), X(-1), \cdots X(-p+1)$, 历史

资料的补充不能减小预测的误差. 当然 (26) 中的系数 b_i 和预测误差一般还要通过形为(23)的方程来求解. 而当 $n_0=1$ 即向前一步进行预测时, (27)就直接给出了预测值, 而预测误差

$$\begin{aligned}\sigma_1^2 &\triangleq E|X(1)-\hat{X}(1)|^2 \\ &=E|X(1)+\sum_{j=1}^{p}a_jX(1-j)|^2 \\ &=E|\xi(1)|^2 \end{aligned} \tag{28}$$

若 $\sigma^2\triangleq E|\xi(1)|^2$, 则 $\sigma_1^2=\sigma^2$.

3. 滑动和模型

若平稳序列 $X_T=\{X(n), n=0, \pm1, \cdots\}$ 可表为如下有限滑动和形式

$$X(n)=\sum_{j=0}^{q}b_j\xi(n-j) \tag{29}$$

其中 $\{\xi(n)\}$ 为白噪声, 并且假定 b_j 满足

$$\sum_{j=0}^{q}b_j\lambda^{q-j}=0 \text{ 的根全在单位圆} |\lambda|<1 \text{ 内} \tag{30}$$

这时由定理 4.4.5,

$$H_n(X)=H_n(\xi)$$
$$\xi(n+k)\perp H_n(X), \qquad k>0$$

所以利用 $\{X(k), k\leqslant0\}$ 对 $X(n_0)$ 的最佳线性预测 $\hat{X}(n_0)$ 为

$$\begin{aligned}\hat{X}(n_0)&=P_{H_0(X)}X(n_0) \\ &=\sum_{j=0}^{q}b_jP_{H_0(X)}\xi(n_0-j) \\ &=\begin{cases}\sum_{j=n_0}^{q}b_j\xi(n_0-j), & n_0\leqslant q \\ 0, & n_0>q\end{cases}\end{aligned} \tag{31}$$

而预测误差为

$$
\begin{aligned}
\sigma_{n_0}^2 &\triangleq E|X(n_0)-\hat{X}(n_0)|^2 \\
&=\sigma^2 \sum_{j=0}^{\min(n_0-1,q)} |b_j|^2
\end{aligned}
\tag{32}
$$

其中 $\sigma^2=E|\xi(0)|^2$. 所以对于 q 阶滑动和过程, 由于它的协方差函数 $B_X(n)$满足

$$
B_X(n)=0, \qquad |n|>q
$$

因此进行 q 步以上的预测时, 历史资料不能提供任何信息, 预测时用不用历史资料是一个样, (31)和(32)提供了预测值和预测误差的公式. 但当 $n_0\leqslant q$ 时, 要用$\{X(n), n\leqslant 0\}$直接表出 $\hat{X}(n_0)$还必须求出$\xi(k)$用$\{X(n), n\leqslant k\}$的表示式, 并代入(31)之中, 引理 4.4.1 表明这种表示是存在的, 但一般说来 $\hat{X}(n_0)$ 必须表示为过去无限项$X(-n)$之和. 这些系数将通过展开某个有理函数来获得(参见下一段对 ARMA 过程预测问题的讨论).

4. ARMA 模型

若平稳序列 $X_T=\{X(n), n=0, \pm 1, \cdots\}$满足如下 ARMA 模型

$$
\sum_{j=0}^{p} a_j X(n-j)=\sum_{j=0}^{q} b_j \xi(n-j) \tag{33}
$$

其中$\{\xi(n)\}$为白噪声, 并且假定$\{a_j\}$, $\{b_j\}$满足

$$
\begin{aligned}
&\sum_{j=0}^{p} a_j \lambda^{p-j}=0 \\
&\sum_{j=0}^{q} b_j \lambda^{q-j}=0
\end{aligned}
\tag{34}
$$

的根全在单位圆$|\lambda|<1$内. 这时由定理 4.5.5

$$
\left.
\begin{aligned}
&H_n(X)=H_n(\xi) \\
&\xi(n+k)\perp H_n(X), k>0
\end{aligned}
\right\}
\tag{35}
$$

并且

$$X(n)=\sum_{k=0}^{\infty}c_k\xi(n-k) \tag{36}$$

这样，利用$\{X(k),k\leqslant 0\}$对$X(n_0)$的最佳线性预测$\hat{X}(n_0)$为

$$\begin{aligned}\hat{X}(n_0)&=P_{H_0(X)}\sum_{k=0}^{\infty}c_k\xi(n_0-k)\\&=\sum_{k=0}^{\infty}c_kP_{H_0(\xi)}\xi(n_0-k)\\&=\sum_{k=n_0}^{\infty}c_k\xi(n_0-k)\end{aligned} \tag{37}$$

预测误差

$$\begin{aligned}\sigma_n^2&\triangleq E|X(n_0)-\hat{X}(n_0)|^2\\&=E|\sum_{j=0}^{\infty}c_j\xi(n-k)-\sum_{j=n_0}^{\infty}c_j\xi(n-k)|^2\\&=\sigma^2\sum_{j=0}^{n_0-1}|c_j|^2\end{aligned} \tag{38}$$

其中$\sigma^2\triangleq E|\xi(0)|^2$，(37)，(38)只是形式上给出了预测问题的解，因为对c_j及$\xi(n)$还未给出它的明显公式．为了给出更明确一点的回答，在频率域上进行相应的讨论是有用的．

若$\xi(n)$有如下谱分解

$$\xi(n)=\int_{-\pi}^{\pi}e^{in\lambda}d\zeta_\xi(\lambda)$$

则由(33)

$$X(n)=\int_{-\pi}^{\pi}e^{in\lambda}C(e^{-i\lambda})d\zeta_\xi(\lambda) \tag{39}$$

其中

$$C(e^{-i\lambda})=\frac{\sum_{j=0}^{q}b_je^{-ij\lambda}}{\sum_{j=0}^{p}a_je^{-ij\lambda}} \tag{40}$$

比较(36)和(39)可得

$$C(e^{-i\lambda})=\sum_{k=0}^{\infty}c_k e^{-ik\lambda}$$

所以 c_k 就是 $C(z)$ 展为幂级数时 z^k 的系数，即

$$c_k=\frac{1}{k!}\frac{d^k}{dz^k}C(z)$$

当然在实用中我们也可利用 $\sum_{j=0}^{p}a_j z^{p-j}=0$ 的根全在圆内这个性质把(40)分解为部分分式而求得 c_k. 若表

$$[e^{in_0\lambda}C(e^{-i\lambda})]_{0-}=\sum_{k=n_0}^{\infty}c_k e^{i(n_0-k)\lambda}$$

则由(37),(39),

$$\begin{aligned}\hat{X}(n_0)&=\int_{-\pi}^{\pi}\left(\sum_{k=n_0}^{\infty}c_k e^{i(n_0-k)\lambda}\right)d\zeta_\xi(\lambda)\\&=\int_{-\pi}^{\pi}\frac{[e^{in_0\lambda}C(e^{-i\lambda})]_{0-}}{C(e^{-i\lambda})}d\zeta_X(\lambda)\end{aligned}\tag{41}$$

其中 $\zeta_X(\lambda)$ 为 X_T 的随机谱函数. 若记

$$\Phi(\lambda)=\frac{[e^{in_0\lambda}C(e^{-i\lambda})]_{0-}}{C(e^{-i\lambda})}$$

那么 $\hat{X}(n_0)=\int_{-\pi}^{\pi}\Phi(\lambda)d\zeta_X(\lambda)$. 在某些工程技术上，只要能给出 $\Phi(\lambda)$,就可设计专门的线性系统，它可近似地实现如下运算. 当输入为 $X(n)$ 时，其输出便是 $\hat{X}(n_0)$，所以对某些应用来说，(41)给出了预测问题解.

5. 一般平稳序列的预测问题

从对ARMA过程预测问题的讨论中可以看出，把平稳序列 $X(n)$ 表为满足(35) 的形为(36)的单边滑动和在解决预测问题中是本质的. 因为只要(35),(36)成立，那么(37),(38)也就成立，至

少在形式上是得到了预测问题的解.所以对一般平稳时间序列,解决它的预测问题的一个途径就是寻求满足(35)的表示式(36). 并设法由协方差函数求出 c_k. 对平稳过程的进一步研究指出①,若平稳时间序列的谱函数为 $F(\lambda)$, $F'(\lambda)$ 表示 $F(\lambda)$ 绝对连续部分导数,则当下列条件成立时,

$$\int_{-\pi}^{\pi} \log F'(\lambda) d\lambda = -\infty$$

必有对任一 n, $X(n) \in H_0(X)$, 即用 $\{X(k), k \leqslant 0\}$ 预测 $X(n)$, $n>0$ 时是无误差的,而当 X_T 存在谱密度 $f(\lambda)$, 并满足

$$\int_{-\pi}^{\pi} \log f(\lambda) d\lambda > -\infty$$

则 X_T 必可表为形为 (36) 的单边滑动和,且可由 $f(\lambda)$ 求出满足(35)的 c_k. 但是这些求 c_k 的公式在实用上要运用仍然是困难的,而且也没有像自回归模型一样的从 $X(n)$, $n \leqslant 0$ 计算 $\hat{X}(n_0)$ 的有效算法.

四、有噪声观测平稳序列的估计

1. 首先我们考虑平稳序列的如下估计问题. 若平稳序列 $X(n)$, $Z(n)$, $V(n)$ 满足下列条件

$$Z(n) = X(n) + V(n) \tag{42}$$

其中 $V(n)$ 与 $X(n)$ 是互不相关的观测噪声. 我们要利用 $\{Z(n)\}$ 的全部观测来对 $X(n_0)$ 进行估计,因为 Z_T 为平稳序列,若 $F_Z(\lambda)$, $\zeta_Z(\lambda)$ 分别表示它的谱函数和随机谱函数,则 Z_T 有如下谱分解

$$Z(n) = \int_{-\pi}^{\pi} e^{in\lambda} d\zeta_Z(\lambda)$$

又 $\hat{X}(n_0)$ 表 Z_T 对 $X(n_0)$ 的最小方差估计, 则因为 $\hat{X}(n_0) \in H(Z)$, 而由(4.2.39)可知,必有 $\varphi(\lambda) \in L_2(dF_Z)$ 使

① 参见郑绍濂等编校"希尔伯脱空间中的平稳序列",上海科技出版社(1962年),第一篇文章.

$$\hat{X}(n_0)=\int_{-\pi}^{\pi}\varphi(\lambda)d\zeta_Z(\lambda)$$

这时表征 $\hat{X}(n_0)$为最佳线性估计的(14)可写为

$$\begin{aligned}&\int_{-\pi}^{\pi}\varphi(\lambda)e^{-in\lambda}dF_Z(\lambda)\\&=E\{\hat{X}(n_0)\overline{Z(n)}\}=E\{X(n_0)\overline{Z(n)}\}\\&=\int_{-\pi}^{\pi}e^{i(n_0-n)\lambda}dF_{XZ}(\lambda),\qquad n=0,\pm1,\cdots\end{aligned}\tag{43}$$

其中 $F_{XZ}(\lambda)$为 X_T, Z_T 的交互谱函数，特别当 X_T, V_T 分别存在谱密度 $f_{XX}(\lambda)$和 $f_{VV}(\lambda)$时，且因它们又互不相关，则由(42)，Z_T 也有谱密度 $f_{ZZ}(\lambda)$，且

$$f_{ZZ}(\lambda)=f_{XX}(\lambda)+f_{VV}(\lambda)$$

$$F_{XZ}(\lambda)=\int_{-\pi}^{\lambda}f_{XX}(\mu)d\mu$$

这时由(43)便可推出

$$\varphi(\lambda)=\frac{e^{in_0\lambda}f_{XX}(\lambda)}{f_{XX}(\lambda)+f_{VV}(\lambda)}\tag{44}$$

$$\hat{X}(n_0)=\int_{-\pi}^{\pi}e^{in_0\lambda}\frac{f_{XX}(\lambda)}{f_{XX}(\lambda)+f_{VV}(\lambda)}d\zeta_Z(\lambda)\tag{45}$$

而估计误差为

$$\begin{aligned}\sigma^2&=E|X(n_0)-\hat{X}(n_0)|^2\\&=E|X(n_0)|^2-E|\hat{X}(n_0)|^2\\&=\int_{-\pi}^{\pi}(f_{XX}(\lambda)-\frac{f_{XX}^2(\lambda)}{f_{XX}(\lambda)+f_{VV}(\lambda)})d\lambda\\&=\int_{-\pi}^{\pi}\frac{f_{XX}(\lambda)f_{VV}(\lambda)}{f_{XX}(\lambda)+f_{VV}(\lambda)}d\lambda\end{aligned}$$

(45) 表明为了由 Z_T 获得对 $X(n_0)$ 的估计，可以设计一个线性系统，它的频率特性是

$$\Psi(\lambda) \triangleq \frac{f_{XX}(\lambda)}{f_{XX}(\lambda)+f_{VV}(\lambda)} \tag{46}$$

这时把 Z_T 输入这一系统时，便可获得 Z_T 对 X_T 的估计. 对于(46)我们可以从一个特殊情况来理解它. 若

$$f_{VV}(\lambda)=0, \quad a\leqslant\lambda\leqslant b$$

$$f_{XX}(\lambda)=0, \quad \lambda<a \text{ 或 } \lambda>b$$

这时按(46)

$$\Psi(\lambda)=\begin{cases}0, & \lambda<a \text{ 或 } \lambda>b \\ 1, & a\leqslant\lambda\leqslant b\end{cases} \tag{47}$$

这就表明当信号 X_T 的谱集中在 $a\leqslant\lambda\leqslant b$，而噪声 V_T 的谱集中在 $\lambda<a$ 或 $\lambda>b$，二者互不相重，这时就可设计以(47)中的 $\Psi(\lambda)$ 为频率特征的线性系统，它让谱集中在 $[a,b]$ 的信号全部通过，而把谱集中在 $\lambda<a$ 或 $\lambda>b$ 的噪声全部压制掉，这与直观想法是完全一致的.

在这里我们是运用了 Z_T 的全部资料来对 $X(n)$ 进行估计，这种估计是无法实时进行的，只能在观测结束后进行估计. 下面我们介绍具有有理谱的平稳序列的滤波问题.

2. 具有有理谱平稳序列的滤波和外推

我们用一个简单的例子来说明具有有理谱密度的平稳序列的滤波和外推问题的频域处理方法. 若观测到的平稳序列 $Z(n)$ 可表为

$$Z(n)=X(n)+V(n)$$

其中 $\{X(n)\}$ 和 $\{V(n)\}$ 是互不相关的平稳序列，它们的谱密度 $f_{XX}(\lambda)$ 和 $f_{VV}(\lambda)$ 分别为

$$f_{XX}(\lambda)=\frac{C_1}{2\pi|e^{i\lambda}-a|^2}, \qquad |a|<1 \tag{48}$$

$$f_{VV}(\lambda)=\frac{C_2}{2\pi}$$

由此，$Z(n)$亦为平稳序列，且它的谱密度为

$$f_{ZZ}(\lambda)=\frac{C_1}{2\pi|e^{i\lambda}-a|^2}+\frac{C_2}{2\pi} \tag{49}$$

此时必可取适当的C和$|b|<1$（参见第4章习题），使$f_{ZZ}(\lambda)$表为

$$f_{ZZ}(\lambda)=\frac{C|e^{i\lambda}-b|^2}{2\pi|e^{i\lambda}-a|^2}, \qquad |a|<1,\ |b|<1 \tag{50}$$

且因为X_T，V_T互不相关，所以X_T和Z_T的交互谱密度f_{XZ}必满足

$$f_{XZ}(\lambda)=f_{XX}(\lambda) \tag{51}$$

我们考虑用$\{Z(n),\ n\leqslant 0\}$对$X(n_0)$，$n_0\geqslant 0$进行估计的问题，由于Z_T以(49)为谱密度，所以$H_0(Z)$中的元素可以表为（参见习题）

$$H_0(Z)=\left\{\int_{-\pi}^{\pi}\varphi(e^{i\lambda})d\zeta_Z(\lambda):\varphi(e^{i\lambda})=\sum_{n=0}^{\infty}c_n e^{-in\lambda},\right.$$

$$\left.\sum_{n=0}^{\infty}|c_n|^2<\infty\right\}$$

其中$\zeta_Z(\lambda)$为Z_T的随机谱函数．若$X(n_0)$的最小方差线性估计为$\hat{X}(n_0)$，则

$$\hat{X}(n_0)=\int_{-\pi}^{\pi}\varphi(e^{i\lambda})d\zeta_Z(\lambda)$$

其中

$$\varphi(e^{i\lambda})=\sum_{n=0}^{\infty}c_n e^{-in\lambda}, \qquad \sum_{n=0}^{\infty}|c_n|^2<\infty \tag{52}$$

又因$\hat{X}(n_0)$是$X(n_0)$的最佳线性估计，由(17)，当$n\leqslant 0$时，

$$\int_{-\pi}^{\pi}\varphi(e^{i\lambda})e^{-in\lambda}f_{ZZ}(\lambda)d\lambda$$

$$=E\{\hat{X}(n_0)\overline{Z(n)}\}=E\{X(n_0)\overline{Z(n)}\}$$

$$=\int_{-\pi}^{\pi} e^{i(n_0-n)\lambda} f_{XZ}(\lambda) d\lambda$$

由(51)，即可推出

$$\int_{-\pi}^{\pi} [\varphi(e^{i\lambda}) f_{ZZ}(\lambda) - e^{in_0\lambda} f_{XX}(\lambda)] e^{-in\lambda} d\lambda = 0,$$

$$n \leqslant 0 \tag{53}$$

所以对

$$\psi(e^{i\lambda}) \triangleq \varphi(e^{i\lambda}) f_{ZZ}(\lambda) - e^{in_0\lambda} f_{XX}(\lambda) \tag{54}$$

由于 f_{ZZ} 和 f_{XX} 都是 $e^{i\lambda}$ 的有理函数，我们不妨从 $e^{i\lambda}$ 的有理函数类中去找 ψ 和 φ，由(53)可考虑

$$\psi(e^{i\lambda}) = \sum_{j=1}^{\infty} a_j e^{ij\lambda}$$

若将(48)，(50)代入(54)可得

$$\frac{C\varphi(e^{i\lambda}) |e^{i\lambda} - b|^2}{|e^{i\lambda} - a|^2} - \frac{C_1 e^{in_0\lambda}}{|e^{i\lambda} - a|^2} = \psi(e^{i\lambda})$$

利用解析延拓唯一性，可得对一切 z 成立

$$\frac{C\varphi(z)(z-b)(1-z\bar{b})}{z}$$

$$= C_1 z^{n_0} + \frac{\psi(z)}{z}(z-a)(1-z\bar{a}) \tag{55}$$

由 $n_0 \geqslant 0$ 及 $\psi(z)$ 是在 $|z|=1$ 上可积的有理函数，上式右端在 $|z| \leqslant 1$ 上是解析的，所以

$$\varphi_1(z) \triangleq \frac{\varphi(z)(z-b)}{z}$$

在 $|z| \leqslant 1$ 上是解析的．另一方面，由(53)，$\varphi_1(z)$ 在 $|z|>1$ 上也是解析的，所以

$$\varphi_1(z) \equiv k, \quad \varphi(z) = \frac{k}{1-bz^{-1}}$$

为求出常数 k，取 $z=a$ 代入(55)，可得

$$k=\varphi_1(a)=\frac{C_1 a^{n_0}}{C(1-a\bar{b})}$$

所以

$$\varphi(e^{i\lambda})=\frac{C_1 a^{n_0}}{C(1-a\bar{b})}\cdot\frac{1}{(1-be^{-i\lambda})}$$

$$\hat{X}(n_0)=\int_{-\pi}^{\pi}\frac{C_1 a^{n_0}}{C(1-a\bar{b})(1-be^{-i\lambda})}d\zeta_Z(\lambda)$$

$$=\frac{C_1 a^{n_0}}{C(1-a\bar{b})}\sum_{n=0}^{\infty}b^n Z(-n)$$

而估计误差为

$$\begin{aligned}\sigma_{n_0}^2&=E|X(n_0)-\hat{X}(n_0)|^2\\&=E|X(n_0)|^2-E|\hat{X}(n_0)|^2\\&=\int_{-\pi}^{\pi}(f_{XX}(\lambda)-|\varphi(\lambda)|^2 f_{ZZ}(\lambda))d\lambda\end{aligned}$$

上面对满足(48)——(50)的平稳序列，介绍了它的预测和滤波问题，虽然这个例子是一个简单的情况，但是这里介绍的寻求估计量频率特征$\varphi(e^{i\lambda})$的方法，对解决一般的有理谱密度情况下已知$\{Z(n),n\leqslant 0\}$估计$X(n_0)$的问题时，同样是适用的，而且也可适用于$n_0\leqslant 0$时内插的情况①.

五、线性估计的递推公式-卡尔曼滤波公式

上面我们虽然用投影定理来说明了线性估计问题的一般解法，但只是对具有有理谱密度的平稳序列具体地讨论了它的线性估计问题，并求出了它的线性估计的频率特征．随着近代控制理论的发展，为了准确地控制某一系统进入最佳状态，必须对系统的状态进行估计．这时由于系统本身是随时间变化的，或者由于干扰本身是时变的，所以把描写系统的状态作为平稳过程来处理就是不够的了．另外在这些系统中，要求估计的状态还是一个多

① 参见 A. M. 雅格龙"平稳随机函数导论"，数学进展，第二卷，3—156(1956)．

维的随机过程，所以必须要对原来考虑的模型加以推广．例如在武器控制系统中，为了有效地进行控制，首先必须对打击目标的状态进行估计，但是描述空间一个飞行目标的位置至少需要三个变量，假若还要考虑它的速度，那就要六个变量，因而用一维随机过程来处理是不够的，此外，由于飞行目标本身的机动性，把飞行目标的轨迹作为平稳过程来处理也是不妥当的，所以必须考虑更为一般的模型．

另一方面，上面我们介绍的线性估计问题的各种解法中，它都只能从时间域或从频率域上求出一个最终表达式，借助这个表达式，可由整个已知的观测量来计算出估计量．随着计算技术的发展和运用计算技术进行实时控制的需要，必须在观测过程中不断对系统状态进行估计，并随着新的观测资料的获得又不断修正这种估计，要是在每次需要进行估计的时候都利用已获得的全部观测值计算一次估计值，那么既要存放大量过去的观测数据，又要化费大量的计算时间，所以必须有适合实时控制需要的线性估计的递推算法，这便是六十年代中发展起来的卡尔曼滤波理论．

在卡尔曼滤波理论中，所考虑的模型是一对随机过程$\{X(n)\}$和$\{Z(n)\}$，若

$$X(n+1)=A(n)X(n)+W(n),\quad n\geqslant 0 \tag{56}$$

$$Z(n)=H(n)X(n)+V(n),\qquad n\geqslant 0 \tag{57}$$

其中(56)又称为系统的**状态方程**，$X(n)$是表示系统状态的随机过程，$X(n)$一般是一个p维的随机过程，$A(n)$为$p\times p$的矩阵，$W(n)$是表示系统本身的噪声，而(57)又称系统的**观测方程或量测方程**，$V(n)$表示观测噪声，引入$H(n)$是由于$X(n)$的分量不是都能直接被观测，往往只能观测到$X(n)$分量的某些线性组合，$Z(n)$一般也是一个多维随机过程，若它的维数为q，则$H(n)$就是一个$q\times p$的矩阵．

对于满足(56)，(57)的X_T和Z_T，我们还假定其中$\{W(n)\}$，$\{V(n)\}$的二阶特征满足下列条件

$$\left.\begin{aligned}E\{W(n)\}=0, E\{W(n)W(m)^{\tau}\}=Q(n)\delta_{nm}\\ E\{V(n)\}=0, E\{V(n)V(m)^{\tau}\}=R(n)\delta_{nm}\end{aligned}\right\} \tag{58}$$

$$E\{W(n)V(m)^{\tau}\}=0$$

$$E\{X(0)V(n)^{\tau}\}=E\{X(0)W(n)^{\tau}\}=0 \tag{59}$$

(58)表明$\{W(n)\}$，$\{V(n)\}$都是不相关的噪声序列，且互不相关，(59)表明$\{W(n)\}\{V(n)\}$与$X(0)$也是不相关的．我们的目的是要由$\{Z(0), Z(1), \cdots, Z(n)\}$作出$X(n)$的最佳线性估计$\hat{X}(n)$．即要使

$$E|X(n)-\hat{X}(n)|^2=\min \tag{60}$$

因为(20)已给出了$\hat{X}(n)$的一般解，所以在这里要介绍的是$\hat{X}(n)$递推算法，为了求得$\hat{X}(n)$及其递推公式，我们仍将使用投影定理．若以$S_n\triangleq$记{由$Z(0)$，$Z(1)$，$\cdots$，$Z(n)$的分量张成的H的子空间}则由定理2

$$\hat{X}(n)=P_{S_n}X(n)\triangleq\begin{bmatrix}P_{S_n}X_1(n)\\ \vdots\\ P_{S_n}X_p(n)\end{bmatrix}$$

以后我们还将多次用到多维随机变量在S_n上的投影，都是指按每个分量进行投影．在解决无噪声平稳序列的外推问题时，满足(35)的$\{\xi(n)\}$是起了重要作用的，实际上在那里$\{\xi(k), k\leqslant n\}$是$H_n(X)$中的一个正交基，它使$P_{H_n}(X)$有方便的形式．现在，我们仍将采用这一想法，在推导有关结论之前，先引入下列记号

$$\hat{X}(n)=P_{S_n}X(n),\quad \hat{X}(n|n-1)=P_{S_{n-1}}X(n) \tag{61}$$

$$\tilde{X}(n)=X(n)-\hat{X}(n),\quad \tilde{X}(n|n-1)=X(n)-\hat{X}(n|n-1) \tag{62}$$

$$P(n)=E\{\tilde{X}(n)\tilde{X}(n)^{\tau}\}$$

$$P(n|n-1)=E\{\tilde{X}(n|n-1)\tilde{X}(n|n-1)^{\tau}\} \tag{63}$$

这里 $\hat{X}(n)$为对 $X(n)$的滤波值, $\hat{X}(n|n-1)$为对 $X(n)$的向前一步预测值, $\tilde{X}(n)$, $\tilde{X}(n|n-1)$分别为滤波和预测误差, 都是 $p\times 1$ 的向量. $P(n)$, $P(n|n-1)$分别为滤波和预测的误差矩阵, 而记号 B^{τ}总表示矩阵或向量 B 的共轭转置.

定理 3 若$\{X(n), n\geqslant 0\}$, $\{Z(n), n\geqslant 0\}$满足(56)——(59), 令

$$\xi(n)=Z(n)-P_{S_{n-1}}Z(n) \tag{64}$$

则$\{\xi(n)\}$满足下列条件

$$S_n=S_{n-1}\oplus H(\xi(n))=S_0\oplus\sum_{i=1}^{n}H(\xi(i)) \tag{65}$$

其中 $H(\xi(n))$表示由 $\xi(n)$的分量张成的H的子空间, 而符号$\oplus$表示相互正交的希尔伯脱空间 S_{n-1} 和 $H(\xi(n))$的线性和. 并且

$$E\{\xi(n)\xi(m)^{\tau}\}$$

$$=\delta_{nm}(H(n)P(n|n-1)H(n)^{\tau}+R(n)) \tag{66}$$

$$E\{X(n)\xi(n)^{\tau}\}=P(n|n-1)H(n)^{\tau} \tag{67}$$

[证明] 由(64)及投影定理

$$\xi(n)\perp S_{n-1}$$

即 $\xi(n)$的每个分量与 S_{n-1} 直交, 所以

$$H(\xi(n))\perp S_{n-1}$$

又由(64), $\xi(n)\in S_n$(指每个分量都成立), 所以

$$H(\xi(n))\oplus S_{n-1}\subset S_n \tag{68}$$

反之, 按(64)

$$Z(n)=\xi(n)+P_{S_{n-1}}Z(n)$$

所以 $Z(n)$的每个分量都属于 $H(\xi(n))\oplus S_{n-1}$, 而 S_n 是由 S_{n-1} 及 $Z(n)$的分量所张成的H的子空间, 所以

$$S_n\subset H(\xi(n))\oplus S_{n-1} \tag{69}$$

联合(68), (69)即得(65). 当 $n>m$ 时

$$\xi(m)\in S_{n-1}, \xi(n)\perp S_{n-1}$$

所以

$$E\{\xi(n)\xi(m)'\}=0, \quad n>m \tag{70}$$

又由(58)，(59)不难推出

$$W(n)\perp S_n, \quad V(n)\perp S_{n-1}$$

$$E\{X(n)V(m)'\}=0$$

利用这一点及(56)，(57)，可推得

$$\begin{aligned}\xi(n)&=Z(n)-P_{S_{n-1}}Z(n)\\&=Z(n)-P_{S_{n-1}}(H(n)X(n)+V(n))\\&=H(n)X(n)+V(n)-H(n)\hat{X}(n|n-1)\\&=H(n)\tilde{X}(n|n-1)+V(n)\end{aligned} \tag{71}$$

因为 $\hat{X}(n|n-1)\in S_{n-1}$，$V(n)\perp S_{n-1}$，所以 $V(n)\perp\hat{X}(n|n-1)$，即

$$E\{\hat{X}(n|n-1)V(n)'\}=0 \tag{72}$$

又 $V(n)\perp X(n)$，所以 $V(n)\perp\tilde{X}(n|n-1)$，即

$$E\{\tilde{X}(n|n-1)V(n)'\}=0 \tag{73}$$

所以

$$\begin{aligned}&E\{\xi(n)\xi(n)'\}\\&=H(n)E\{\tilde{X}(n|n-1)\tilde{X}(n|n-1)'\}H'(n)+E\{V(n)V(n)'\}\\&=H(n)P(n|n-1)H(n)'+R(n)\end{aligned} \tag{74}$$

联合(70)，(74)即得(66)，又因

$$X(n)=\tilde{X}(n|n-1)+\hat{X}(n|n-1) \tag{75}$$

由(71)，(75)及(72)，(73)可得

$$\begin{aligned}&E\{X(n)\xi(n)'\}\\&=E\{\tilde{X}(n|n-1)\tilde{X}(n|n-1)'\}H(n)'\\&\quad+E\{\tilde{X}(n|n-1)V(n)'\}\\&\quad+E\{\hat{X}(n|n-1)\tilde{X}(n|n-1)'\}H(n)'\\&\quad+E\{\hat{X}(n|n-1)V(n)'\}\end{aligned}$$

$$=E\{\tilde{X}(n|n-1)\tilde{X}(n|n-1)'\}H(n)'$$
$$=P(n|n-1)H(n)'$$

即(67)成立，定理证毕.

事实上，S_n 是观测进行到 n 为止，我们可能用以对 $X(n)$ 进行估计的随机变量全体，是代表到 n 为止观测到的信息. 随着 n 的增大，已知的信息也越多，即 $S_n \supset S_{n-1}$. 而(64)及 $\xi(n) \perp S_{n-1}$ 表明 $\xi(n)$ 是代表在时刻 n 所获得的新信息. 所以通常又称 $\{\xi(n)\}$ 为 $\{Z(n)\}$ 所对应的**新息序列**. 若直到时刻 $n-1$ 为止，我们用了 S_{n-1} 中知识对 $X(n)$ 进行了最佳估计，那么到时刻 n，我们只需用新息 $\xi(n)$ 来修正这一估计就可得到对 $X(n)$ 的最佳估计，这就是获得线性估计递推算法的基本出发点，严格地叙述这一结论的是下列定理.

定理 4 若 $\{X(n), n\geqslant 0\}$，$\{Z(n), n\geqslant 0\}$ 满足(56)——(59)，且 $R(n)$ 为满秩阵，则由 $\{Z(0), Z(1), \cdots, Z(n)\}$ 对 $X(n)$ 的滤波值 $\hat{X}(n)$ 满足下列递推关系

$$\hat{X}(n)=A(n-1)\hat{X}(n-1)+K(n)\xi(n)$$
$$=A(n-1)\hat{X}(n-1)$$
$$+K(n)(Z(n)-H(n)A(n-1)\hat{X}(n-1)) \qquad (76)$$

其中 $\xi(n)$ 由(64)规定，

$$K(n)=P(n|n-1)H(n)'(H(n)P(n|n-1)H(n)'+R(n))^{-1}$$
$$=P(n)H(n)'R(n)^{-1} \qquad (77)$$

而 $P(n)$ 和 $P(n|n-1)$ 又满足下列递推关系式

$$P(n|n-1)=A(n-1)P(n-1)A(n-1)'+Q(n-1) \qquad (78)$$

$$P(n)=(I-K(n)H(n))P(n|n-1)\cdot(I-K(n)H(n))'$$
$$+K(n)R(n)K(n)'$$
$$=(I-K(n)H(n))P(n|n-1)$$
$$=(P(n|n-1)^{-1}+H(n)'R(n)^{-1}H(n))^{-1} \qquad (79)$$

在最后一个式子中还假定一切 $Q(n)$ 是满秩的。

［证明］ 由(65)

$$\begin{aligned}\hat{X}(n)&=P_{S_n}X(n)=P_{S_{n-1}\oplus H(\xi(n))}X(n)\\&=P_{S_{n-1}}X(n)+P_{H(\xi(n))}X(n)\\&=P_{S_{n-1}}(A(n-1)X(n-1)+W(n-1))+K(n)\xi(n)\\&=A(n-1)\hat{X}(n-1)+K(n)\xi(n)\end{aligned}\tag{80}$$

而其中 $K(n)$ 由(66),(67)及 $R(n)$ 为满秩的，可得

$$\begin{aligned}K(n)&=E\{X(n)\xi(n)'\}(E\{\xi(n)\xi(n)'\})^{-1}\\&=P(n|n-1)H(n)'(H(n)P(n|n-1)H(n)'+R(n))^{-1}\\&=(P(n|n-1)^{-1}+H(n)'R(n)^{-1}H(n))^{-1}H(n)'R(n)^{-1}\end{aligned}\tag{81}$$

上式最后一步当 $Q(n)$ 为非奇异时可用右乘

$$H(n)P(n|n-1)H(n)'+R(n)$$

直接验证，又由(71)

$$\begin{aligned}\xi(n)&=Z(n)-P_{S_{n-1}}H(n)X(n)\\&=Z(n)-P_{S_{n-1}}(H(n)A(n-1)X(n-1)+H(n)W(n-1))\\&=Z(n)-H(n)A(n-1)\hat{X}(n-1)\end{aligned}$$

所以(76)成立。由于

$$\begin{aligned}\hat{X}(n|n-1)&=P_{S_{n-1}}X(n)\\&=P_{S_{n-1}}(A(n-1)X(n-1)+W(n-1))\\&=A(n-1)\hat{X}(n-1)\end{aligned}$$

所以按(62)有

$$\tilde{X}(n|n-1)=A(n-1)\tilde{X}(n-1)+W(n-1)$$

利用 $W(n-1)$ 与 $X(n-1)$ 和 S_{n-1} 的正交性，可以得到

$$\begin{aligned}P(n|n-1)&=E\{\tilde{X}(n|n-1)\tilde{X}(n|n-1)'\}\\&=A(n-1)P(n-1)A(n-1)'+Q(n-1)\end{aligned}$$

即(78)成立．所以当 $Q(n-1)$ 满秩时，$P(n|n-1)$ 也非奇异，推导(81)的运算是合理的．另一方面，由(80)

$$\hat{X}(n)=\hat{X}(n|n-1)+K(n)\xi(n)$$

所以由(63)及(71)可推出

$$\begin{aligned}\widetilde{X}(n)&=\widetilde{X}(n|n-1)-K(n)\cdot(H(n)\widetilde{X}(n|n-1)+V(n))\\&=(I-K(n)H(n))\widetilde{X}(n|n-1)-K(n)V(n)\end{aligned}$$

由(81)，

$$\begin{aligned}I-K(n)H(n)&=(P(n|n-1)^{-1}+H(n)'R(n)^{-1}H(n))^{-1}\\&\quad(P(n|n-1)^{-1}+H(n)'R(n)^{-1}H(n)-H(n)'R(n)^{-1}H(n))\\&=(P(n|n-1)+H(n)'R(n)^{-1}H(n))^{-1}P(n|n-1)^{-1}\end{aligned}\tag{82}$$

所以由上式及(81)，

$$\begin{aligned}P(n)&=E\{\widetilde{X}(n)\widetilde{X}(n)'\}\\&=(I-K(n)H(n))P(n|n-1)(I-K(n)H(n))'\\&\qquad+K(n)R(n)K(n)'\\&=(P(n|n-1)^{-1}+H(n)'R(n)^{-1}H(n))^{-1}\\&\qquad\cdot(I-H(n)'K(n)'+H(n)'K(n)')\\&=(P(n|n-1)^{-1}+H(n)'R(n)^{-1}H(n))^{-1}\end{aligned}\tag{83}$$

将此式代入(81)即得(77)，而又由(82)及上式

$$I-K(n)H(n)=P(n)P(n|n-1)^{-1}$$

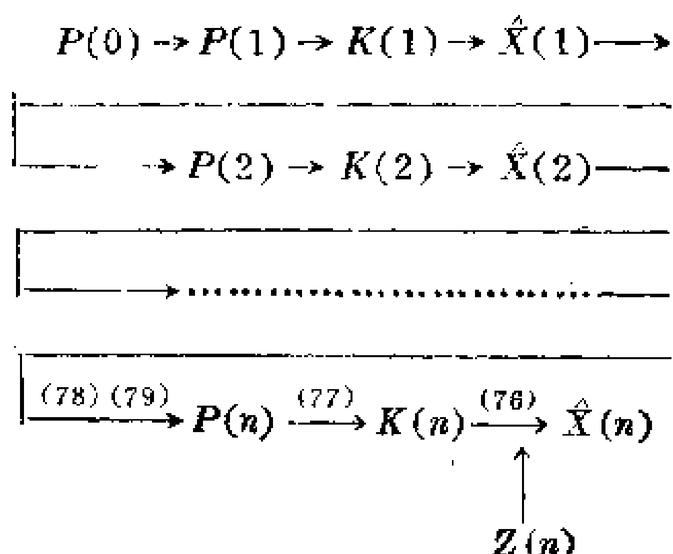
所以

$$P(n)=(I-K(n)H(n))P(n|n-1)\tag{84}$$

由(83)，(84)即得(79)，定理证毕

(76)——(79)就是计算滤波值的递推公式，又称为**卡尔曼滤波公式**．在实际运用这些公式时，首先要确定使用这些递推公式的初态，即求出 $\hat{X}(0)$ 和 $P(0)$，然后按下列次序随着观测资料的增

长逐步递推地计算 $\hat{X}(n)$:



假若注意到(77)——(79)的 $K(n)$ 和 $P(n)$ 与观测值 $Z(n)$ 是无关的，那么为了在实时处理中加快计算速度，也可以事先或由其它计算机把各个 $K(n)$ 和 $P(n)$ 都算好，这时获得新的观测 $Z(n)$ 时，就可立即由(76)算出 $X(n)$ 的估计值 $\hat{X}(n)$，这也是一种计算 $X(n)$ 的办法。对于滤波值的递推公式，我们可以这样来解释它，为方便起见可认为 $X(n)$，$Z(n)$ 是一维的。因为 $\hat{X}(n-1)$ 是到 $n-1$ 为止对 $X(n-1)$ 的最佳估计，由于 $X(n)$ 本身的递推关系(56)，所以 $\hat{X}(n|n-1)=A(n-1)\hat{X}(n-1)$，但到时刻 n 获得了新的观测 $Z(n)$，它为我们提供了新息 $\xi(n)$，这时我们有可能利用它来修正对 $X(n)$ 的估计. 这一修正就在 $A(n-1)X(n-1)$ 之外加上一项 $K(n)\xi(n)$，$K(n)$ 表示加权因子，当 $\xi(n)=Z(n)-\hat{X}(n|n-1)$ 较大时，即原来估计 $\hat{X}(n|n-1)$ 与新获得的对 $X(n)$ 的观测 $Z(n)$ 相差较大，修正要大一些，若 $\xi(n)=0$，可不必修正。另一方面，从 $K(n)$ 的表示式(77)可看出，若观测噪声方差 $R(n)$ 较小，则 $K(n)$ 要大一些，这是因为此时 $Z(n)$ 代表 $X(n)$ 较好，$Z(n)-\hat{X}(n|n-1)$ 能较好地反映 $X(n)-\hat{X}(n|n-1)$ 间的偏差，因而修正的权 $K(n)$ 可大一些，而由(78)，(79)可见，当 $Q(n)$ 增大时，$K(n)$ 亦随之增大，这是因为 $Q(n)$

的增大表示 $X(n)$ 对 $A(n-1)X(n-1)$ 的变化较大，所以 $\hat{X}(n)$ 对 $A(n-1)\hat{X}(n-1)$ 的修正亦宜加大，这样，滤波的递推公式从直观上来看也是合理的.

利用滤波值 $\hat{X}(n)$，不难求出预测值如下

$$P_{S_n}X(n+m)=A(n+m-1)\cdots A(n+1)A(n)\hat{X}(n)$$

最后，我们来说明滤波递推算法的一个重要性质.

定理 5 若 $\{X(n), n\geqslant 0\}$ 和 $\{Z(n), n\geqslant 0\}$ 满足 (56)——(59)，且 $H(n)\equiv 1$，$\xi(n)$ 和 $\hat{X}(n)$ 分别由 (64)，(76) 给出，若对 $X(0)$ 的初始滤波值 $\hat{X}(0)$ 满足等式

$$E\{\hat{X}(0)\}=E\{X(0)\} \tag{85}$$

则对一切 n 成立等式

$$\begin{aligned} E\{\hat{X}(n)\}&=E\{\hat{X}(n|n-1)\} \\ &=E\{X(n)\} \end{aligned} \tag{86}$$

$$E\{\xi(n)\}=0 \tag{87}$$

[证明] 我们对 n 用归纳法证明 (86)，由 (56)，(57)

$$E\{X(1)\}=A(0)E\{X(0)\}$$

$$E\{Z(1)\}=E\{X(1)\}=A(0)E\{X(0)\}$$

所以由 (76)，

$$\begin{aligned} E\{\hat{X}(1)\}&=A(0)E\{\hat{X}(0)\} \\ &\quad +K(1)(E\{Z(1)\}-A(0)E\{\hat{X}(0)\}) \\ &=A(0)E\{X(0)\} \\ &\quad +K(1)(A(0)E\{X(0)\}-A(0)E\{X(0)\}) \\ &=A(0)E\{X(0)\}=E\{X(1)\} \end{aligned}$$

又由 $\hat{X}(n|n-1)=A(n-1)\hat{X}(n-1)$，所以

$$\begin{aligned} E\{\hat{X}(1|0)\}&=A(0)E\{\hat{X}(0)\} \\ &=A(0)E\{X(0)\} \\ &=E\{X(1)\} \end{aligned}$$

若对 $n=l$, (86)成立, 则当 $n=l+1$ 时, 由(56), (57),

$$E\{X(l+1)\}=A(l)E\{X(l)\}$$

$$E\{Z(l+1)\}=E\{X(l+1)\}$$

由(76)

$$\begin{aligned}E\{\hat{X}(l+1)\}&=A(l)E\{\hat{X}(l)\}\\&+K(l+1)(E\{Z(l+1)\}-A(l)E\{\hat{X}(l)\})\\&=A(l)E\{X(l)\}\\&=E\{X(l+1)\}\end{aligned}$$

又因 $\hat{X}(l+1|l)=A(l)\hat{X}(l)$, 所以

$$\begin{aligned}E\{\hat{X}(l+1|l)\}&=A(l)E\{\hat{X}(l)\}\\&=A(l)E\{X(l)\}\\&=E\{X(l+1)\}\end{aligned}$$

因而(86)成立. 又由(71),

$$\xi(n)=Z(n)-\hat{X}(n|n-1)$$

所以

$$\begin{aligned}E\{\xi(n)\}&=E\{Z(n)\}-E\{\hat{X}(n|n-1)\}\\&=E\{X(n)\}-E\{X(n)\}\\&=0\end{aligned}$$

(85)也成立, 即定理为真.

通常还把满足(86)的滤波和预测称为无偏的, 所以只要初始估计满足无偏性条件(85), 那么按卡尔曼滤波公式获得的估计还是无偏的.

§2. 线性模型的均值估计

一、模型与例

在上一节, 我们介绍了时间序列的预测和滤波问题. 在那里,

都是假定时间序列的某些统计特征(例如均值，协方差)为已知的前提下来讨论问题的，对于如何由时间序列观测到的样本来推断它的某些统计特征，这是本节和以后几节要解决的问题. 我们并不在最一般的情况下来讨论这一问题，因为随机过程的统计问题远比随机变量分布的统计问题困难得多，至今也只对某些有用的特殊情况有一些可行的做法. 在这一节里，我们假定被观测的时间序列 $Z(n)$ 有如下的形式

$$\left.\begin{array}{l} Z(n)=m(n)+X(n) \\ E\{X(n)\}=0, E\{X(n)\overline{X(m)}\}=B(n-m) \end{array}\right\} \tag{1}$$

其中 $m(n)$ 为非随机的序列，为 $Z(n)$ 的均值部分，$X(n)$ 是均值为零的平稳时间序列，是 $Z(n)$ 的随机变化部分，又称为余差.

这是一个十分有用的模型，因为在很多测量或观测中，所得到的时间序列它一方面有某种决定性的变化趋向，例如化学反应中的温度、压力、浓度等等的决定性的变化，气象、水文资料中温度、气压、水位的周期变化，地震记录中的地震波等等，上述 $m(n)$ 就表示这一决定性变化的. 另一方面，在测量或观测的记录中，又不可避免地混有随机干扰，包括记录误差、系统噪声和其它种种干扰，上述 $X(n)$ 就表示这种干扰，这里把 $X(n)$ 假定为零均值的平稳时间序列也能满足大部分实际需要，它当然包括了 $X(n)$ 为白噪声这一重要特例，有时，我们感兴趣的就是某个平稳时间序列的统计性质，它不一定是干扰或噪声，例如在海浪波高分析时，海浪波高本身可假定为平稳过程，它的统计特征是我们感兴趣的对象，这种情况也可包括在上述模型之内，只需令 $m(n)\equiv m$ 表示平均波高，$X(n)$ 表示波高相对于平均波高的变化，$Z(n)$ 就是我们能观测到的波高变化.

为了便于统计分析，模型(1)中的 $m(n)$ 也不是任意形式的，我们假定它是具有下列形式的：

$$m(n)=\sum_{\nu=1}^{l}\alpha_\nu\varphi_\nu(n) \tag{2}$$

其中 $\varphi_\nu(n)$ 可表为

$$\varphi_\nu(n)=n^{k_\nu}\cdot\cos\lambda_\nu n$$
$$n^{k_\nu}\cdot\sin\lambda_\nu n \text{ 或 } n^{k_\nu}\cdot e^{i\lambda_\nu n} \tag{3}$$

我们规定 $m(n)$ 具有这种形式是因为它可以包括许多实际需要的情况，而且又便于对它进行统计分析，因为在(3)中，若取 $k_\nu=0$，$\varphi_\nu(n)=\cos n\lambda_\nu$ 或 $\sin n\lambda_\nu$ 就可表示均值的周期变化，它适用于气象，水文，天文中很多情况，当 $\lambda_\nu=0$，$\varphi_\nu(n)=n^{k_\nu}$ 则可表示均值具有不同增长速度的趋势变化，所以我们以它为讨论的对象就可适应多种需要，而这里所用的方法也可帮助我们处理更一般的模型.

对于形为(1)，(2)的模型，在第二部分第四章的线性估计理论中也曾遇到过，不过那里的 $X(n)$ 取为白噪声序列. 现在我们称 $\varphi_\nu(n)$，$1\leqslant\nu\leqslant l$ 为回归变量，α_ν 又称为回归系数. 而模型(1)，(2)就称为**具有平稳余差的线性回归模型**，简称**线性模型**.

二、回归系数的估计问题

当(2)中的 k_ν，λ_ν 为已知时，回归变量 $\varphi_\nu(n)$ 不含未知参数. 这时依据 $Z(1),\cdots,Z(N)$ 对回归系数 α_ν 的估计问题在第二部分第四章中曾介绍用最小二乘法来求解，即令

$$\sum_{n=1}^{N}\left|Z(n)-\sum_{\nu=1}^{l}\alpha_\nu\varphi_\nu(n)\right|^2=\min \tag{4}$$

来求解 α_ν，$1\leqslant\nu\leqslant l$. (4)又可归为下列 α_ν 的线性方程组，

$$\sum_{j=1}^{l}\alpha_j\left(\sum_{n=1}^{N}\overline{\varphi_i(n)}\varphi_j(n)\right)=\sum_{n=1}^{N}\overline{\varphi_i(n)}Z(n) \tag{5}$$

这样求得的 $\hat{\alpha}_\nu$，$1\leqslant\nu\leqslant l$ 又称为 α_ν，$1\leqslant\nu\leqslant l$ 的**最小二乘方估计**，简写为 LS 估计.

在处理这类问题时，采用矩阵表示是方便的，若记

$$
Z=\begin{bmatrix} Z(1) \\ \vdots \\ Z(N) \end{bmatrix}, \qquad X=\begin{bmatrix} X(1) \\ \vdots \\ X(N) \end{bmatrix}
$$

$$
\alpha=\begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_l \end{bmatrix}, \qquad \Phi=\begin{bmatrix} \varphi_1(1)\cdots\varphi_l(1) \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ \varphi_1(N)\cdots\varphi_l(N) \end{bmatrix}
$$

则

$$
Z=\Phi\alpha+X
$$

(4)又可写为

$$
(Z-\Phi\alpha)'(Z-\Phi\alpha)=\min
$$

其中 B' 表矩阵 B 的共轭转置阵，而(5)可写为

$$
\Phi'\Phi\alpha=\Phi'Z \tag{6}
$$

所以当 $\Phi'\Phi$ 满秩时(相当于 l 个回归变量 $\varphi_\nu(n)$, $1\leqslant\nu\leqslant l$ 是线性无关的)，α 的 LS 估计 $\hat{\alpha}_{LS}$ 可表为

$$
\hat{\alpha}_{LS}=(\Phi'\Phi)^{-1}\Phi'Z \tag{7}
$$

这是在线性估计理论中已介绍的做法，那里曾假定了 X 为(正态)白噪声. 现在 X 为平稳余差，即 $X(n)$之间有统计联系. 在目前的条件下，LS 估计的做法仍然是可行的，但它的统计性质又怎样呢，下面我们来讨论这一问题.

定理 1 $\hat{\alpha}_{LS}$ 是 α 的线性无偏估计.

[证明] 由(6)，$\hat{\alpha}_{LS}$ 是 Z 分量的线性函数，至于无偏性，因为

$$
\begin{aligned}
E\{\hat{\alpha}_{LS}\}&=E\{(\Phi'\Phi)^{-1}\Phi'Z\}\\
&=(\Phi'\Phi)^{-1}\Phi'E\{Z\}\\
&=(\Phi'\Phi)^{-1}\Phi'\Phi\alpha=\alpha
\end{aligned}
$$

定理 2 若 $B=(B(i-j))$ 表示由 $X(n)$ 的协方差函数构成的 $N\times N$ 矩阵，又 B, $\Phi'B^{-1}\Phi$ 都满秩，则(i)由

$$
(Z-\Phi\alpha)'B^{-1}(Z-\Phi\alpha)=\min \tag{8}
$$

求得的 α 为

$$\hat{\alpha}_M=(\Phi' B^{-1}\Phi)^{-1}\Phi' B^{-1}Z \tag{9}$$

(ii) $\hat{\alpha}_M$ 为 α 的线性无偏估计；

(iii) 若 $\hat{\beta}$ 亦为 α 的线性无偏估计，$\Sigma_{\hat{\beta}}$ 表 $\hat{\beta}$ 的协方差阵，则 $\Sigma_{\hat{\beta}}-\Sigma_{\hat{\alpha}_M}$ 为非负定阵.

[证明] (i)记

$$\begin{aligned} Z-\Phi\alpha &= (Z-\Phi\hat{\alpha}_M)+(\Phi\hat{\alpha}_M-\Phi\alpha) \\ &= a_1+a_2 \end{aligned}$$

则

$$\begin{aligned} &(Z-\Phi\alpha)' B^{-1}(Z-\Phi\alpha) \\ &= a_1' B^{-1}a_1+a_2' B^{-1}a_1+a_1' B^{-1}a_2+a_2' B^{-1}a_2 \end{aligned}$$

而

$$\begin{aligned} a_2' B^{-1}a_1 &= (\Phi\hat{\alpha}_M-\Phi\alpha)' B^{-1}(Z-\Phi\hat{\alpha}_M) \\ &= (\hat{\alpha}_M-\alpha)'\Phi' B^{-1}(I-\Phi(\Phi' B^{-1}\Phi)^{-1}\Phi' B^{-1})Z \\ &= (\hat{\alpha}_M-\alpha)'(\Phi' B^{-1}-\Phi' B^{-1})Z=0 \end{aligned}$$

同理 $a_1' B^{-1}a_2=0$, 所以

$$\begin{aligned} &(Z-\Phi\alpha)' B^{-1}(Z-\Phi\alpha)=a_1 B^{-1}a_1+a_2 B^{-1}a_2 \\ &= (Z-\Phi\hat{\alpha}_M)' B^{-1}(Z-\Phi\hat{\alpha}_M)+(\hat{\alpha}_M-\alpha)'\Phi' B^{-1}\Phi(\hat{\alpha}_M-\alpha) \end{aligned}$$

因此当 $\alpha=\hat{\alpha}_M$ 时，(8)达到最小值.

(ii) 显然，$\hat{\alpha}_M$ 是 Z 分量的线性函数，又

$$\begin{aligned} E\{\hat{\alpha}_M\} &= E\{(\Phi' B^{-1}\Phi)^{-1}\Phi' B^{-1}Z\} \\ &= (\Phi' B^{-1}\Phi)^{-1}\Phi' B^{-1}E\{Z\} \\ &= (\Phi' B^{-1}\Phi)^{-1}\Phi' B^{-1}\Phi\alpha=\alpha \end{aligned}$$

即 $\hat{\alpha}_M$ 是无偏的.

(iii) 若

$$C=\begin{bmatrix} C_{11} & \cdots & C_{1l} \\ \cdots & \cdots & \cdots \\ C_{N1} & \cdots & C_{Nl} \end{bmatrix}$$

$\hat{\beta}=C'Z$ 为 α 的无偏估计，则由无偏性

$$\alpha=E\{\hat{\beta}\}=E\{C'Z\}=C'\Phi\alpha$$

所以

$$C'\Phi=I$$

而

$$\begin{aligned}\hat{\beta}-\alpha&=C'Z-\alpha\\&=C'\Phi\alpha+C'X-\alpha=C'X\end{aligned}$$

所以

$$\begin{aligned}\Sigma_{\hat{\beta}}&=E(\hat{\beta}-\alpha)(\hat{\beta}-\alpha)'\\&=E\{C'XX'C\}=C'BC\end{aligned}$$

对于 $\hat{\alpha}_M$, $C'=(\Phi'B^{-1}\Phi)^{-1}\Phi'B^{-1}$，所以

$$\begin{aligned}\Sigma_{\hat{\alpha}_M}&=(\Phi'B^{-1}\Phi)^{-1}\Phi'B^{-1}BB^{-1}\Phi(\Phi'B^{-1}\Phi)^{-1}\\&=(\Phi'B^{-1}\Phi)^{-1}\end{aligned} \tag{10}$$

今考虑矩阵 $A'=C'-(\Phi'B^{-1}\Phi)^{-1}\Phi'B^{-1}$，则 $A'BA$ 为非负定阵，另一方面，利用 $C'\Phi=I$ 可推得

$$\begin{aligned}A'BA&=(C'-(\Phi'B^{-1}\Phi)^{-1}\Phi'B^{-1})\\&\cdot B(C-B^{-1}\Phi(\Phi'B\Phi)^{-1})\\&=C'BC-(\Phi'B^{-1}\Phi)^{-1}\\&=\Sigma_{\hat{\beta}}-\Sigma_{\hat{\alpha}_M}\end{aligned}$$

所以 $\Sigma_{\hat{\beta}}-\Sigma_{\hat{\alpha}_M}$ 为非负定的，定理证毕.

由于非负定矩阵的对角线元素都是非负的，而 $\Sigma_{\hat{\beta}}-\Sigma_{\hat{\alpha}_M}$的对角线元素就是 $D(\hat{\beta}_1)-D(\hat{\alpha}_M)$, $D(\hat{\beta}_2)-D(\hat{\alpha}_{M_2})$, $\cdots$, $D(\hat{\beta}_l)-D(\hat{\alpha}_{M_l})$，所以定理 2 告诉我们对 $\hat{\alpha}_M$ 的每个分量 $\hat{\alpha}_{M\nu}$

$$E\{\hat{\alpha}_{M\nu}\}=\alpha_\nu$$

$$D\{\hat{\alpha}_{M\nu}\}\leqslant D\{\hat{\beta}_\nu\}$$

即 $\hat{\alpha}_M$ 的每个分量都是最小方差线性无偏估计，作为 l 维向量 α 的估计 $\hat{\alpha}_M$，定理 2 表明它是线性无偏估计中具有最小方差阵的。

当 $\{X(n)\}$ 为白噪声时，

$$B=\sigma^2 I$$

$$\begin{aligned}\hat{\alpha}_M&=(\Phi'\sigma^{-2}I\Phi)^{-1}\Phi'\sigma^{-2}IZ\\&=(\Phi'\Phi)^{-1}\Phi'Z=\hat{\alpha}_{LS}\end{aligned}$$

所以 $\hat{\alpha}_{LS}$ 是 X 为白噪声时的最小方差无偏估计，而在一般的相关余差情况下，由定理 2 就有

$$D(\hat{\alpha}_{LS\nu})\geqslant D(\hat{\alpha}_{M\nu}),\quad 1\leqslant\nu\leqslant l$$

所以从估计量的方差来看，$\hat{\alpha}_M$ 比 $\hat{\alpha}_{LS}$ 好，但是由(9)可知，$\hat{\alpha}_M$ 只有当 B 已知的情况下才可求得，即必须知道 X_T 的协方差函数才能求得 $\hat{\alpha}_M$，而且为了计算 $\hat{\alpha}_M$，必须要算出 B^{-1} 或 $B^{-1}\Phi$，而估计量 $\hat{\alpha}_{LS}$ 却不需要了解 X_T 的协方差函数 B，它的计算也比较方便，这正是 $\hat{\alpha}_{LS}$ 吸引人的地方．那么 $\hat{\alpha}_{LS}$ 和 $\hat{\alpha}_M$ 的方差差别有多大呢，能否采用 $\hat{\alpha}_{LS}$ 而"基本上"与 $\hat{\alpha}_M$ 有相同的方差呢．下面的定理将回答这一问题．

定理 3 若 $\{Z(n)\}$ 满足线性模型(1)，(2)，余差 $\{X(n)\}$ 存在谱密度 $f(\lambda)$，它在 $\lambda_1,\cdots,\lambda_l$ 连续且不为零，又以 $\hat{\alpha}_{LS}^{(N)},\hat{\alpha}_M^{(N)}$ 分别表示用 $Z(1),\cdots,Z(N)$ 作出的 α 的 LS 估计和最小方差线性无偏估计，则 $\hat{\alpha}_{LS}^{(N)},\hat{\alpha}_M^{(N)}$ 都是(均方)一致的，即

$$\lim_{N\to\infty}D(\hat{\alpha}_{LS\nu}^{(N)})=\lim_{N\to\infty}D(\hat{\alpha}_{M\nu}^{(N)})=0,\ 1\leqslant\nu\leqslant l \tag{11}$$

并且

$$\lim_{N\to\infty}\frac{D(\hat{\alpha}_{M\nu}^{(N)})}{D(\hat{\alpha}_{LS\nu}^{(N)})}=1 \tag{12}$$

[证明] 我们并不对一般的情况进行证明，因为需要进行较多的计算，我们只对 $l=1,\varphi(n)=e^{in\lambda_1}$ 且 $X(n)$ 的谱密度 $f(\lambda)$ 为

正连续函数的情况加以证明，以此来说明证明的思路①。

对 $l=1,\varphi(n)=e^{in\lambda_0}$，则

$$\hat{\alpha}_{LS}^{(N)}=\left(\sum_{n=1}^{N}|e^{in\lambda_0}|^2\right)^{-1}\left(\sum_{n=1}^{N}e^{-in\lambda_0}Z(n)\right)$$

$$=\frac{1}{N}\sum_{n=1}^{N}e^{-in\lambda_0}X(n)+\alpha$$

利用 $X(n)$ 的谱分解式可得

$$\hat{\alpha}_{LS}^{(N)}-\alpha=\frac{1}{N}\int_{-\pi}^{\pi}\sum_{n=1}^{N}e^{in(\lambda-\lambda_0)}d\xi_X(\lambda)$$

$$=\frac{1}{N}\int_{-\pi}^{\pi}e^{i\frac{N+1}{2}(\lambda-\lambda_0)}\frac{\sin\frac{N}{2}(\lambda-\lambda_0)}{\sin\frac{1}{2}(\lambda-\lambda_0)}d\xi_X(\lambda)$$

其中 $\xi_X(\lambda)$ 为 X_T 的随机谱函数，所以

$$ND(\hat{\alpha}_{LS}^{(N)})=NE|\hat{\alpha}_{LS}^{(N)}-\alpha|^2$$

$$=\int_{-\pi}^{\pi}\frac{\sin^2\frac{N}{2}(\lambda-\lambda_0)}{N\sin^2\frac{1}{2}(\lambda-\lambda_0)}f(\lambda)d\lambda$$

当 $f(\lambda)$ 在 λ_0 连续时，利用富利埃分析中的费耶(Fejer)定理可知

$$\lim_{N\to\infty}ND(\hat{\alpha}_{LS}^{(N)})=2\pi f(\lambda_0) \tag{13}$$

所以(11)成立.

另一方面，若 $\hat{\alpha}=\sum_{n=1}^{N}b(n)Z(n)$ 为 α 的无偏估计，则

$$\alpha=E\{\alpha\}=\sum_{n=1}^{N}b(n)E\{Z(n)\}$$

① 一般情况下的证明可参见王寿仁，“格子点上随机场的回归系数的估计问题”，数学学报，第8卷，210——221页(1958).

$$=\sum_{n=1}^{N} b(n)\varphi(n)\cdot\alpha$$

所以
$$\sum_{n=1}^{N} b(n)\varphi(n)=1$$

而
$$\begin{aligned}\hat{\alpha}&=\sum_{n=1}^{n} b(n)Z(n)\\&=\alpha\sum_{n=1}^{N} b(n)\varphi(n)+\sum_{n=1}^{N} b(n)X(n)\\&=\alpha+\sum_{n=1}^{N} b(n)X(n)\end{aligned}$$

所以

$$\begin{aligned}D(\hat{\alpha})=E|\hat{\alpha}-\alpha|^2=E\left|\sum_{n=1}^{N} b(n)X(n)\right|^2\\&=E\left|\int_{-\pi}^{\pi}\left(\sum_{n=1}^{N} b(n)e^{in\lambda}\right)d\xi_X(\lambda)\right|^2\\&=\int_{-\pi}^{\pi}\left|\sum_{n=1}^{N} b(n)e^{in\lambda}\right|^2 f(\lambda)d\lambda\end{aligned}$$

若取$h(\lambda)=\left(\sum_{n=1}^{N} b(n)e^{in\lambda}\right)\sqrt{f(\lambda)}$,

$$g(\lambda)=\left(\frac{1}{2\pi}\sum_{n=1}^{N}\overline{\varphi(n)}e^{in\lambda}\right)\frac{1}{\sqrt{f(\lambda)}},\text{ 则}$$

$$D(\alpha)=\int_{-\pi}^{\pi}|h(\lambda)|^2 d\lambda$$

$$\int_{-\pi}^{\pi} h(\lambda)\overline{g(\lambda)}d\lambda$$

$$=\int_{-\pi}^{\pi}\left(\sum_{n=1}^{N} b(n)e^{in\lambda}\right)\sqrt{f(\lambda)}\left(\frac{1}{2\pi}\sum_{n=1}^{N}\varphi(n)e^{-in\lambda}\right)\frac{1}{\sqrt{f(\lambda)}}d\lambda$$

$$=\sum_{n=1}^{N}b(n)\varphi(n)=1$$

另一方面由许瓦兹不等式

$$\left|\int_{-\pi}^{\pi}h(\lambda)\overline{g(\lambda)}d\lambda\right|^2\leqslant\int_{-\pi}^{\pi}|h(\lambda)|^2\int_{-\pi}^{\pi}|g(\lambda)|^2d\lambda$$

所以

$$D(\hat{\alpha})=\int_{-\pi}^{\pi}|h(\lambda)|^2d\lambda\geqslant\frac{\left|\int_{-\pi}^{\pi}h(\lambda)\overline{g(\lambda)}d\lambda\right|^2}{\int_{-\pi}^{\pi}|g(\lambda)|^2d\lambda}$$

$$=\left(\frac{1}{4\pi^2}\int_{-\pi}^{\pi}\left|\sum_{n=1}^{N}\varphi(n)e^{-in\lambda}\right|^2\frac{d\lambda}{f(\lambda)}\right)^{-1}$$

$$=\left(\frac{1}{4\pi^2}\int_{-\pi}^{\pi}\left|\sum_{n=1}^{N}e^{in(\lambda-\lambda_0)}\right|^2\frac{1}{f(\lambda)}d\lambda\right)^{-1}$$

$$=\left(\frac{1}{4\pi^2}\int_{-\pi}^{\pi}\frac{\sin^2\frac{N}{2}(\lambda-\lambda_0)}{\sin\frac{1}{2}(\lambda-\lambda_0)}\cdot\frac{d\lambda}{f(\lambda)}\right)^{-1}$$

特别对$\alpha_M^{(N)}$，上式也应成立，所以有

$$1\geqslant\frac{D(\hat{\alpha}_M^{(N)})}{D(\hat{\alpha}_{LS}^{(N)})}\geqslant\frac{4\pi^2}{ND(\hat{\alpha}_{LS}^{(N)})}\left(\int_{-\pi}^{\pi}\frac{\sin^2\frac{N}{2}(\lambda-\lambda_0)}{N\sin^2\frac{1}{2}(\lambda-\lambda_0)}\cdot\frac{d\lambda}{f(\lambda)}\right)^{-1}\tag{14}$$

也由费耶定理可知，

$$\lim_{N\to\infty}\int_{-\pi}^{\pi}\frac{\sin^2\frac{N}{2}(\lambda-\lambda_0)}{N\sin^2\frac{1}{2}(\lambda-\lambda_0)}\cdot\frac{d\lambda}{f(\lambda)}=\frac{2\pi}{f(\lambda_0)}\tag{15}$$

所以由(13)，(15)可推出，当$N\to\infty$时，(14)右端也趋于1，这样就有

$$\lim_{N\to\infty}\frac{D(\hat{\alpha}_M^{(N)})}{D(\hat{\alpha}_{\mathrm{LS}}^{(N)})}=1$$

定理证毕.

上述定理告诉我们,模型(1),(2)中的回归系数 α_ν, $1\leqslant\nu\leqslant l$ 在具有连续正谱密度的平稳余差的条件下,它的最小二乘方估计 $\hat{\alpha}_{\mathrm{LS}}$ 仍是一个良好的估计,一方面它不要求了解余差的协方差函数,计算也较方便,另一方面,随着观测时间的增大,$\hat{\alpha}_{\mathrm{LS}}$ 与最小方差线性无偏估计又是渐近等效的,即二者方差之比趋于1. 所以它为我们提供了 α 的一个切实可用的估计.

对于最小二乘方估计,我们也可以像卡尔曼滤波一样,获得它的递推计算公式. 对此,只需运用§1的结果即可,因为对线性模型(6)可按卡尔曼滤波所考虑的模型写为如下形式

$$\left.\begin{aligned}&\alpha=I\alpha\\&Z(n)=\Phi(n)\alpha+X(n)\end{aligned}\right\}\tag{16}$$

其中 α 为 $l\times1$ 向量表回归系数,$Z(n)$为一维序列,$\Phi(n)=(\varphi_1(n),\cdots,\varphi_l(n))$为 $1\times l$ 向量,$X(n)$为一维余差,在此我们假定它为白噪声序列

$$E\{X(n)\overline{X(m)}\}=\delta_{nm}\cdot\sigma^2$$

在模型(1.56)——(1.59)中,若取$X(n)\equiv\alpha$, $W(n)\equiv0$, $V(n)=X(n)$, $Q(n)\equiv0$, $R(n)\equiv\sigma^2$, 且不难验证(16)也是符合(1.56)——(1.59)这种形式的. 则由定理1.5可知,只要对α的初始估计 $\alpha(n_0)$是无偏的,那么由(1.76)——(1.79)求得的$\hat{\alpha}(n)$也是无偏的. 若初始估计是最小方差无偏的,那么由(1.76)——(1.79)求得的 $\alpha(n)$也是最小方差无偏的,而它们的递推关系由(1.76)——(1.79)按(16)可化简为

$$\hat{\alpha}(n)=\hat{\alpha}(n-1)+K(n)(Z(n)-\Phi(n)\hat{\alpha}(n-1))\tag{17}$$

$$K(n)=\frac{1}{\sigma^2}P(n)\Phi(n)' \tag{18}$$

$$P(n)=\left(P(n-1)^{-1}+\frac{1}{\sigma^2}\Phi(n)'\Phi(n)\right)^{-1} \tag{19}$$

若记 $S(n)=\frac{1}{\sigma^2}P(n)$，则(18)，(19)可化简为

$$K(n)=S(n)\Phi(n)' \tag{20}$$

$$S(n)^{-1}=S(n-1)^{-1}+\Phi(n)'\Phi(n) \tag{21}$$

(17)，(20)，(21)就成为计算 α 的最小方差无偏估计的递推关系式，当然为了做到这一点，还必须有一个合适的初值．一般可以在积累了一定数量的观测值后(例如当 $n\geqslant l$ 时) 直接用(7)来估计 α 得到 $\hat{\alpha}(n)$ 的初态，而这时

$$S(n)=\left(\frac{1}{\sigma^2}\Sigma_{\hat{\alpha}(n)}\right)^{-1}=(\Phi'\Phi)^{-1}$$

以此作为初态，由(17)，(20)，(21)就可递推地求得 α 的最小方差无偏估计，这是余差 $X(n)$ 为白噪声的情形，当 $X(n)$ 为一般的平稳序列时，上述做法仍然可以进行，但是由定理 2 的结论可知，这样求得的 $\hat{\alpha}(n)$ 是 α 的最小二乘方估计，而并不是最小方差无偏估计，不过因为它不需要余差协方差的信息，上述计算还是可行的，又由定理 3 的渐近等效性，所以这种算法仍然是很有实用价值的．

上面讨论了 α 估计量的二阶统计特征问题，关于 α 估计量的分布问题，可以这样来考虑，若 X_T 为正态过程时，则因为 $\hat{\alpha}_{LS}$，$\hat{\alpha}_M$ 都是线性估计，所以它们都是正态分布的，对一般的平稳过程，我们若不对它的分布附加特殊的要求，当然很难得出关于 $\hat{\alpha}_{LS}$，$\hat{\alpha}_M$ 分布的任何结论，但若 X_T 本身为"弱相依"的平稳序列，那么利用弱相依平稳序列的中心极限定理，仍可得到关于 $\hat{\alpha}_{LS}$ 渐近分布的结论，例如

定理 4 若 $Z_T=\{Z(n), n=0, 1, 2, \cdots\}$ 满足(1)——(3)，又

余差 X_T 有连续正谱密度，且满足下列条件之一：

$$\text{(i)}\ X(n)=\sum_{k=-\infty}^{\infty} a_k\xi(n-k),\ \sum_{k=-\infty}^{\infty}|a_k|^2<\infty \tag{22}$$

其中$\{\xi(n)\}$为相互独立同分布随机变量序列；

(ii) $X(n)$ (或 $\xi(n)$) 为有限相依的，即当 $n_i-m_i\geq k$ 时，$(X(n_1),X(n_2),\cdots,X(n_k))$和$(X(m_1),\cdots,X(m_l))$相互独立.

则当N趋于无穷时，$\hat{\alpha}_{LS}^{(N)},\alpha_M^{(N)}$ 都是渐近正态的，特别对其每个分量$\hat{\alpha}_\nu^{(N)}$有

$$\lim_{N\to\infty}P\left(\frac{\alpha_\nu^{(N)}-\alpha_\nu}{\sqrt{D(\alpha_{LS\nu}^{(N)})}}<x\right)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-\frac{u^2}{2}}\,du \tag{23}$$

[证明] 我们就 $l=1,\varphi(n)=e^{in\lambda_\nu}$，即一个回归变量及 X_T 满足(i)的情况下对 $\hat{\alpha}_{LS}$ 来加以证明，以此来说明证明这一结论的一种方法。若 $l=1$，则

$$Z(n)=\alpha e^{in\lambda_0}+X(n)$$

$$\hat{\alpha}_{LS}^{(N)}=\alpha+\frac{1}{N}\sum_{n=1}^{N}X(n)e^{-in\lambda_0}$$

所以

$$\begin{aligned}\sqrt{N}(\hat{\alpha}_{LS}^{(N)}-\alpha)&=\frac{1}{\sqrt{N}}\sum_{n=1}^{N}X(n)e^{-in\lambda_0}\\&=\frac{1}{\sqrt{N}}\sum_{n=1}^{N}\left(\sum_{k=-\infty}^{\infty}a_{n-k}\xi(k)\right)e^{-in\lambda_0}\\&=\frac{1}{\sqrt{N}}\sum_{k=-\infty}^{\infty}\left(\sum_{n=1}^{N}a_{n-k}e^{-in\lambda_0}\right)\xi(k)\\&=\sum_{k=-\infty}^{\infty}b_N(k)\xi(k)\end{aligned} \tag{24}$$

---

① 对于这一定理的证明及推广可参看E. J. Hannen, Multivariate Time Series, 1970. 及 В. И. Ибрагимов, Ю. В. Линник, "Независимые и стационарно связанные. величины", 1965.

其中

$$b_N(k)=\frac{1}{\sqrt{N}}\sum_{n=1}^{N}a_{n-k}e^{-in\lambda_0}$$

对于这一 $b_N(k)$ 我们证明下列两点:

(i) 存在 j_N, 使 $\sum\limits_{|k|\geqslant j_N}|b_N(k)|^2<\frac{1}{N}$; (25)

(ii) $\limsup\limits_{N\to\infty}\sup\limits_{k}|b_N(k)|=0$. (26)

因为由(24)

$$\sum_{k=-\infty}^{\infty}|b_N(k)|^2=ND(\hat{\alpha}_{LS}^{(N)})<\infty$$

所以(25)成立. 另一方面也利用许伐兹不等式

$$\begin{aligned}|b_N(k)|&=\frac{1}{\sqrt{N}}\left|\sum_{n=1}^{N}a_{n-k}e^{-in\lambda_0}\right|\\&=\frac{1}{\sqrt{N}}\left|\sum_{n=1-k}^{N-k}a_ne^{-in\lambda_0}\right|\\&\leqslant\frac{1}{\sqrt{N}}\sum_{|n|\leqslant M}|a_n|+\Big(\sum_{|n|\geqslant M}|a_n|^2\Big)^{\frac{1}{2}}\end{aligned}$$

其中 M 为任一整数, 所以

$$\limsup_{N\to\infty}\sup_{k}|b_N(k)|\leqslant\Big(\sum_{|n|\geqslant M}|a_n|^2\Big)^{\frac{1}{2}}$$

因为 M 是任意的, 由(22)在上式中令 M 趋于无穷即得(26), 利用 $b_N(k)$ 的这两点性质, 我们可按(24)把 $\hat{\alpha}_{LS}$ 表为

$$\begin{aligned}\sqrt{N}(\hat{\alpha}_{LS}^{(N)}-\alpha)&=\sum_{k=-\infty}^{\infty}b_N(k)\xi(k)\\&=\sum_{k=-j_N+1}^{j_N}\eta_{Nk}\end{aligned}\tag{27}$$

其中

$$\eta_{Nk} = \begin{cases} b_N(k)\xi(k), & |k| < j_N \\ \sum\limits_{|k| \geq j_N} b_N(k)\xi(k), & k = j_N \end{cases}$$

它对不同的 k 是相互独立的. 我们来验证(27)中的和式是满足林德伯格条件的. 若

$$B_N^2 = D\{\sqrt{N}(\hat{\alpha}_{LS}^{(N)} - \alpha)\}$$

则由(13)

$$\lim_{N \to \infty} B_N^2 = 2\pi f(\lambda_0) \tag{28}$$

若 $D\{\xi(n)\} = \sigma^2$, 则由(25)可得

$$\frac{1}{B_N^2} \sum_{k=-j_N+1}^{j_N} \int_{|x| \geq \varepsilon B_N} x^2 dF_{\eta_{Nk}}(x)$$

$$\leq \frac{1}{B_N^2}\left(\frac{\sigma^2}{N} + \sum_{k=-j_N+1}^{j_N-1} b_N^2(k) \int_{|x| > \frac{\varepsilon B_N}{|b_N(k)|}} x^2 dF_{\xi(k)}(x)\right)$$

$$\leq \frac{1}{B_N^2}\left(\frac{\sigma^2}{N} + \left(\sum_{k=-\infty}^{\infty} b_N^2(k)\right) \int_{|x| > \frac{\varepsilon B_N}{\sup_k |b_N(k)|}} x^2 dF_{\xi(0)}(x)\right)$$

因此由(26)及(28)可得

$$\lim_{N \to \infty} \frac{1}{B_N^2} \sum_{k=-j_N+1}^{j_N} \int_{|x| \geq \varepsilon B_N} x^2 dF_{\eta_{Nk}}(x) = 0$$

所以$\{\eta_{Nk}\}$是满足林德伯格条件的. 这样由第二册 §5.5 的结果可知对 $\hat{\alpha}_{LS}^{(N)}$, (23)成立. 定理证毕.

至此我们得到了回归系数估计的渐近分布, 利用它可以在大子样情况下作出 α 的渐近置信区间或进行检验.

三、λ_ν 的估计问题

在许多实际问题处理形为(1)——(3)的模型的统计分析问题时, 不仅回归系数 α_ν 是未知的, 而且回归变量 $\varphi_\nu(n)$ 中也含有某些未知参数. 例如在飞机、汽车发动机性能分析时, 发动机的振动数

据就可表为

$$Z(n)=\sum_{\nu=1}^{l}\alpha_\nu e^{i\lambda_\nu n}+X(n) \tag{29}$$

其中 $X(n)$ 为测量误差或其它随机影响，λ_ν 表示发动机的某一振动频率，$|\alpha_\nu|$ 表示相应的频率振动的振幅，它们都是需要由 $Z(n)$ 估计的．了解发动机在不同条件下运转时振动的频率和不同频率振动的幅度和相位是衡量发动机的重要标准，也是改进发动机的依据．在天文、气象、水文资料的分析中也有类似的情况．例如有人把逐年地球公转一次时间的数据用上述模型来描述（在我们日常的习惯中，公转周期是不变的，约 $365\frac{1}{4}$ 天，但精密的天文测量表明它是年年不一样的，而掌握这种变化对研究其他天文、地球物理现象是有用的）．因为天文工作者的经验表明这个 $Z(n)$ 包含了很多隐蔽的周期变化，而且这些周期变化还是和其它天文现象的周期变化密切相关的，所以 $Z(n)$ 也可用(29)来表示．$\alpha_\nu e^{i\lambda_\nu n}$ 表示不同周期的变化项，而 $X(n)$ 则表示其它一些次要的随机影响和测量误差．在此 λ_ν 也是未知的，对它的估计也是十分重要的，只有了解了 λ_ν，才有可能估计 α_ν 并进一步对 $Z(n)$ 的变化进行预测。

为了推导 λ_ν 的估计，我们考虑最简单的 $l=1$ 的情况，记 $\lambda_1=\lambda_0, \alpha_1=\alpha$，我们仍有矩阵表示

$$Z=\Phi\alpha+X, \qquad \Phi=\begin{bmatrix} e^{i\lambda_0} \\ e^{i\lambda_0 2} \\ \vdots \\ e^{i\lambda_0 N} \end{bmatrix}$$

不妨先来看 λ_0, α 的最小二乘方估计，

$$(Z-\Phi\alpha)^\tau(Z-\Phi\alpha)=\min$$

若记 $\hat{\alpha}=(\Phi^\tau\Phi)^{-1}\Phi^\tau Z$ 为 α 的最小二乘方估计(λ_0 已知时)，则

$$(Z-\Phi\alpha)^\tau(Z-\Phi\alpha)$$

$$=[(Z-\Phi\hat{\alpha})+(\Phi\alpha-\Phi\hat{\alpha})]'[(Z-\Phi\hat{\alpha})+(\Phi\hat{\alpha}-\Phi\alpha)]$$
$$=(Z-\Phi\hat{\alpha})'(Z-\Phi\hat{\alpha})+(\alpha-\hat{\alpha})'\Phi'\Phi(\hat{\alpha}-\alpha)$$

所以

$$\min_{\alpha,\lambda_0}(Z-\Phi\alpha)'(Z-\Phi\alpha)$$
$$=\min_{\lambda_0}(Z-\Phi\hat{\alpha})'(Z-\Phi\hat{\alpha})+\min_{\alpha}(\hat{\alpha}-\alpha)'\Phi'\Phi(\hat{\alpha}-\alpha)$$
$$=\min_{\lambda_0}(Z-\Phi\hat{\alpha})'(Z-\Phi\hat{\alpha})$$

又因为 $\hat{\alpha}=(\Phi'\Phi)^{-1}\Phi'Z$，所以

$$(Z-\Phi\hat{\alpha})'(Z-\Phi\hat{\alpha})$$
$$=(Z-\Phi(\Phi'\Phi)^{-1}\Phi'Z)'(Z-\Phi(\Phi'\Phi)^{-1}\Phi'Z)$$
$$=Z'Z-Z'\Phi(\Phi'\Phi)^{-1}\Phi'Z$$

由于 $\Phi'\Phi=N$，$\Phi'Z=\sum_{n=1}^{N}e^{-in\lambda_0}Z(n)$，所以

$$\min_{\lambda_0}(Z-\Phi\hat{\alpha})'(Z-\Phi\hat{\alpha})$$
$$=Z'Z-\max_{\lambda_0}|\Phi'Z|^2/N$$
$$=Z'Z-\max_{\lambda_0}\left(\frac{1}{N}\left|\sum_{n=1}^{N}e^{-in\lambda_0}Z(n)\right|^2\right)$$

若记

$$I_{ZN}(\lambda)=\frac{1}{2\pi N}\left|\sum_{n=1}^{N}e^{-in\lambda}Z(n)\right|^2 \tag{30}$$

则 λ_0 的最小二乘方估计就是使 $I_{ZN}(\lambda)=\max$ 的 $\hat{\lambda}_0$. 在时间序列分析中，由(30)规定的 $I_{ZN}(\lambda)$ 作为 λ 的函数又称为 Z 的**周期图**，因此 λ_0 的最小二乘方估计就是使周期图达到最大的 λ 值，所以也称为**周期图极大估计**.

上面我们是从最小二乘法来获得 λ_0 的估计量，即周期图极大估计. 当 $\{X(n)\}$ 为正态白噪声时，它就是 λ 的最大似然估计，可以期望它有各种最优性质. 当 $\{X(n)\}$ 为一般平稳余差时，我们也

可以进行如下的粗略分析. 由于

$$\sum_{n=1}^{N} e^{-in\lambda} Z(n)$$

$$= \sum_{n=1}^{N} e^{-in\lambda} X(n) + \alpha \sum_{n=1}^{N} e^{in(\lambda_0-\lambda)}$$

$$= \int_{-\pi}^{\pi} \sum_{n=1}^{N} e^{in(\mu-\lambda)} d\xi_X(\mu) + C_2(\lambda)$$

$$= C_1(\lambda) + C_2(\lambda)$$

若$\{X(n)\}$有连续谱密度 $f(\lambda)$, 则

$$\frac{1}{N} E\{|C_1(\lambda)|^2\}$$

$$= \frac{1}{N} E\left|\int_{-\pi}^{\pi} \sum_{n=1}^{N} e^{in(\mu-\lambda)} d\xi_X(\mu)\right|^2$$

$$= \int_{-\pi}^{\pi} \frac{\sin^2 \frac{N}{2}(\mu-\lambda)}{N \sin^2 \frac{1}{2}(\mu-\lambda)} f(\mu) d\mu$$

利用富利埃分析中的费耶定理

$$\lim_{N\to\infty} \frac{1}{N} E|C_1(\lambda)|^2 = 2\pi f(\lambda) \tag{31}$$

所以可以粗略认为$\frac{1}{\sqrt{N}}|C_1(\lambda)|$是有界的, 而$\frac{1}{N}C_1(\lambda)$是趋于零的. 另一方面

$$C_2(\lambda) = \begin{cases} \alpha \cdot N, & \lambda = \lambda_0 \\ \alpha \cdot e^{i\frac{N+1}{2}(\lambda_0-\lambda)} \dfrac{\sin \frac{N}{2}(\lambda-\lambda_0)}{\sin \frac{1}{2}(\lambda-\lambda_0)}, & \lambda \neq \lambda_0 \end{cases}$$

所以

$$\lim_{N\to\infty} \frac{1}{N^2} |C_2(\lambda)|^2 = \begin{cases} \alpha^2, & \lambda = \lambda_0 \\ 0, & \lambda \neq \lambda_0 \end{cases} \tag{32}$$

因此 $$I_{ZN}(\lambda)=\frac{1}{2\pi N}|C_1(\lambda)+C_2(\lambda)|^2$$
$$=\frac{1}{2\pi N}(|C_1(\lambda)|^2$$
$$+2\mathrm{Re}(C_1(\lambda)C_2(\lambda))+|C_2(\lambda)|^2)$$

由(31)及(32)，近似地有

$$\lim_{N\to\infty}I_{ZN}(\lambda)=\begin{cases}\infty, & \lambda=\lambda_0\\ 0, & \lambda\neq\lambda_0\end{cases}$$

所以当N较大时，$I_{ZN}(\lambda)$就在λ_0附近形成一个突起的峰值，而$\lambda\neq\lambda_0$时，$I_{ZN}(\lambda)$的数值远较λ_0附近的数值为小，因此我们利用使周期图$I_{ZN}(\lambda)$达到最大值的λ作为λ_0的估计是十分有利的，作为上述分析的精确化是下列定理.

定理 5 若随机过程$\{Z(n)\}$可表为

$$Z(n)=\alpha e^{-in\lambda_0}+X(n)$$

其中$\{X(n)\}$为平稳正态序列，其谱密度$f(\lambda)$在λ_0连续，则

(i) $I_{ZN}(\lambda)$必在$[-\pi,\pi]$上某$\hat{\lambda}_{0N}$达到最大值；

(ii) 对每个$e>0$，成立

$$\lim_{N\to\infty}P(N|\hat{\lambda}_{0N}-\lambda_0|>e)=0$$

(iii) $$\lim_{N\to\infty}P(N^{-3/2}(\hat{\lambda}_{0N}-\lambda_0)<x)$$
$$=\frac{1}{\sqrt{2\pi}\,\sigma}\int_{-\infty}^{x}e^{-\frac{u^2}{2\sigma^2}}du$$

其中$\sigma^2=\frac{12\pi}{|\alpha|^2}f(\lambda_0)$.

这个定理的证明要进行较长的论证，我们不在这里介绍了. 可以指出，它对相当广泛的一类弱相依平稳过程$\{X(n)\}$都是成立的. 定理 5 的(i)表示周期图极大估计一定存在. (ii)表示这个估计是一致性估计，即随着N增大，估计值$\hat{\lambda}_{0N}$总是越来越靠近λ_0，而且

这种接近速度也比较快，即 $\hat{\lambda}_{0N}-\lambda_0=o\left(\frac{1}{N}\right)$. (iii) 更表明 $\hat{\lambda}_{0N}$ 还是渐近正态的，所以定理 5 表明 λ_0 的周期图极大估计是一个优良的估计.

当 $Z(n)$ 具有多个周期回归变量时，它可表为

$$Z(n)=\sum_{\nu=1}^{l}\alpha_\nu e^{i\lambda_\nu n}+X(n)$$

这时仍可进行类似于上面的分析. 对 λ_1, …, λ_l 的估计应取为 $I_{ZN}(\lambda)$ 的 l 个最大的相对极大值. 当我们事先具有关于 λ_ν 的某个信息，例如 $|\lambda_i-\lambda_j|>\delta$ 时，则其估计 $\hat{\lambda}_\nu$ 亦应取得满足 $|\hat{\lambda}_i-\hat{\lambda}_j|>\delta$. 对于多个回归变量的情况，定理 5 仍然是成立的，即 $I_{ZN}(\lambda)$ 一定存在 l 个相对极大值（其间隔 $>\delta$），周期图极大估计是一致的，还是渐近正态的，所以对多个周期回归变量来说，周期图极大估计仍然是一个优良的估计.

在许多实际问题中，究竟有多少个回归变量是预先不知道的，在分析发动机振动和天文资料时都是如此. 我们采用周期图极大估计，往往可以得到很多个相对极值，究竟应该取多少个回归变量才合适呢. 从原则上讲，这个问题也可用如下的参数假设检验来处理：当我们从周期图极大估计中获得了 λ_ν 的估计后，$e^{i\lambda_\nu n}$ 是不是真的 $Z(n)$ 所隐含的周期分量呢，这一点是用检验 $e^{i\lambda_\nu n}$ 所对应的回归系数 α_ν 是否显著的不为零来实现，因为当 $e^{i\lambda_\nu n}$ 是 $Z(n)$ 真正隐含的周期分量，则 α_ν 应不为零，而当我们人为硬性加入 $e^{i\lambda_\nu n}$ 作为回归分量时，那么 α_ν 不能显著地异于零，但要确切地建立起这个问题的检验法又与 $\{X(n)\}$ 的功率谱特性有关，而这正是待估计的. 所以实用的方法还是半经验性的，要建立严密的检验法是困难的.

§3. 余差的协方差和谱的估计

在上一节里，我们对如下的线性模型

$$Z(n)=m(n)+X(n)=\sum_{\nu=1}^{l}a_{\nu}\varphi_{\nu}(n)+X(n) \tag{1}$$

$$E\{X(n)\}=0$$

$$E\{X(n)\overline{X(m)}\}=B(n-m)$$

考虑了它的均值$m(n)$的估计问题，在这一节里，我们将针对模型(1)讨论它的余差平稳序列$\{X(n)\}$的协方差函数和谱密度的估计问题。对不同的时间序列，由于它的用途不同，在分析它的余差时也往往假定为具有不同形式的模型。当我们需要对时间序列进行预测、控制时，往往喜欢把$\{X(n)\}$假定为满足自回归模型或更一般的自回归滑动和混合模型，即假定它满足某个形式已知而含有若干个未知参数的模型。这样的时间序列要用于预测和控制是较为方便的，而且当观测到资料不多时也便于估计出这些参数，在这种情况下往往就要从协方差函数估计量出发来求出未知参数的估计量。但是在另一些用途中，由于功率谱的物理含义，我们感兴趣的对象就是功率谱本身。例如海浪的波谱分析，汽车本身振动的波谱分析，这时不论是否先进行协方差函数的估计，但所有讨论都将集中在功率谱上，即在频率域上进行分析。它们都是对$X(n)$的二阶统计特征或与其有关的参数进行估计，所以这一节我们将先介绍基本的协方差函数的估计问题和功率谱密度的估计问题，而在下一节再展开参数模型的估计问题的讨论。

由于要对$X(n)$的协方差函数或功率谱等二阶特征进行估计，所以往往要采用$Z(n)$的二次函数来构造估计量，而当我们涉及这一估计量的方差时，就自然会遇到$Z(n)$的四阶矩的计算。这时，只假定余差$\{X(n)\}$为一般平稳时间序列是不够的，必须附加关于$\{X(n)\}$四阶矩的限制，为简便起见，在这一节我们都假定$\{X(n)\}$为实的平稳正态时间序列，虽然本节的一些结果可推广到更一般的四阶平稳时间序列上，一些具体估计方法也可用于一般的平稳

时间序列，但正态性的假定可简化计算，便于我们掌握所用的方法，而在应用中又不是太大的限制.

若$\{X(n)\}$为实的零均值正态序列，那么利用特征函数与矩的关系，不难推出下列等式

$$\begin{aligned}&E\{X(n_1)X(n_2)X(n_3)X(n_4)\}\\&\quad=E\{X(n_1)X(n_2)\}E\{X(n_3)X(n_4)\}\\&\qquad+E\{X(n_1)X(n_3)\}E\{X(n_2)X(n_4)\}\\&\qquad+E\{X(n_1)X(n_4)\}E\{X(n_2)X(n_3)\}\end{aligned}\tag{2}$$

特别的 $$E\{X(n)\}^4=3(E\{X(n)\}^2)^2\tag{3}$$

这是我们以后在推导估计量方差时经常要用到的等式.

一、协方差函数的估计

为了引出$\{X(n)\}$协方差函数$B(n)$的估计，我们不妨先假定$m(n)\equiv 0$，这时，$Z(n)=X(n)$，我们是可以直接观测到$X(n)$的，由于它是平稳正态的，所以对每个ν，$\{X(n+\nu)X(n), n=0, \pm 1, \cdots\}$也是平稳正态的，而且

$$B(\nu)=E\{X(n+\nu)X(n)\}$$

依照§4.3介绍的遍历性定理，自然会想到用$\{X(n+\nu)X(n)\}$的时间平均来作为它的统计均值$B(\nu)=E\{X(n+\nu)X(n)\}$的估计. 所以记

$$\hat{B}^*(\nu)=\frac{1}{N-\nu}\sum_{n=1}^{N-\nu}X(n+\nu)X(n),\qquad \nu\geqslant 0$$

用它来估计$B(\nu)$. 要注意，当我们获得观测值$\{X(1),\cdots,X(N)\}$，用它来计算$\hat{B}^*(\nu)$时，由于需要$X(n+\nu)$，所以上式中对n的求和范围只能由1到$N-\nu$. 当ν可正可负时，我们就用如下表示

$$\hat{B}^*(\nu)=\frac{1}{N-|\nu|}\sum_{n=1}^{N-|\nu|}X(n+|\nu|)X(n)\tag{4}$$

在时间序列分析中，常要用$\hat{B}^*(\nu)$来计算功率谱估计，为了保证谱

密度估计为正的，因此常希望协方差函数的估计 $\hat{B}(\nu)$是一个正定序列，所以更经常采用的是下列估计

$$\hat{B}(\nu)=\frac{1}{N}\sum_{n=1}^{N-|\nu|}X(n+|\nu|)X(n) \tag{5}$$

上面是在 $m(n)\equiv 0$ 的条件下来讨论的，当 $m(n)\not\equiv 0$ 时，我们便无法直接观测到 $X(n)$，这时我们可先求出 $m(n)$的估计．若 $\hat{\alpha}_\nu$ 表示 α_ν 的最小二乘方估计或其它比它有效的估计，

$$\hat{m}(n)=\sum_{\nu=1}^{l}\hat{\alpha}_\nu\varphi_\nu(n) \tag{6}$$

$$\begin{aligned}\hat{X}(n)&=Z(n)-\hat{m}(n)\\&=X(n)-(\hat{m}(n)-m(n))\end{aligned} \tag{7}$$

我们就可从$\{\hat{X}(n)\}$出发来估计 $B(n)$，象(5)中 $\hat{B}(\nu)$一样，取

$$\hat{B}_{\hat{X}}(\nu)=\frac{1}{N}\sum_{n=1}^{N-|\nu|}\hat{X}(n+|\nu|)\hat{X}(n) \tag{8}$$

在我们开始介绍 $\hat{B}(\nu)$的性质时，先证明下列定理．

定理 1 若$\{Z(n)\}$满足(1)，且$\{X(n)\}$为具有连续谱密度的平稳正态序列，回归变量 $\varphi_\nu(n)$取为(2.3)的形式，则存在与 N，ν 无关的常数 K，使

$$N^2E|\hat{B}_{\hat{X}}(\nu)-\hat{B}(\nu)|^2\leqslant K \tag{9}$$

[证明] 由(5)，(7)及(8)：

$$\begin{aligned}&\hat{B}_{\hat{X}}(\nu)-\hat{B}(\nu)\\&=\frac{1}{N}\sum_{n=1}^{N-|\nu|}\{[X(n+|\nu|)-(\hat{m}(n+|\nu|)-m(n+|\nu|))]\\&\qquad\cdot[X(n)-(\hat{m}(n)-m(n))]-X(n+|\nu|)X(n)\}\\&=\frac{1}{N}\sum_{n=1}^{N-|\nu|}(\hat{m}(n+|\nu|)-m(n+|\nu|))(\hat{m}(n)-m(n))\end{aligned}$$

$$-\frac{1}{N}\sum_{n=1}^{N-|\nu|}(\hat{m}(n+|\nu|)-m(n+|\nu|))X(n)$$

$$-\frac{1}{N}\sum_{n=1}^{N-|\nu|}(\hat{m}(n)-m(n))X(n+|\nu|) \tag{10}$$

为了推导方便,我们只对 $m(n)=\alpha e^{in\lambda_1}$, $\hat{\alpha}=\hat{\alpha}_{LS}$ 的情况加以证明,对一般情况也可类似地进行. 在这情况下

$$\hat{m}(n)=\hat{\alpha}_{LS}e^{i\lambda_0 n}$$
$$=\frac{1}{N}\left(\sum_{k=1}^{N}Z(k)e^{-ik\lambda_0}\right)e^{in\lambda_0}$$

$$\hat{m}(n)-m(n)$$
$$=\frac{1}{N}\left(\sum_{k=1}^{N}X(k)e^{-ik\lambda_0}\right)e^{in\lambda_0}$$

$$\sum_{n=1}^{N-|\nu|}(\hat{m}(n)-m(n))X(n+|\nu|)$$
$$=\frac{1}{N}\left(\sum_{k=1}^{N}X(k)e^{-ik\lambda_0}\right)\left(\sum_{n=1}^{N-|\nu|}X(n+|\nu|)e^{in\lambda_0}\right)$$
$$=\frac{1}{N}Z_1Z_2$$

利用(2)

$$E\left|\sum_{n=1}^{N-|\nu|}(\hat{m}(n)-m(n))X(n+|\nu|)\right|^2$$
$$=\frac{1}{N^2}EZ_1Z_2\bar{Z}_1\bar{Z}_2$$
$$=\frac{1}{N^2}(E|Z_1|^2E|Z_2|^2+E\{Z_1Z_2\}E\{\bar{Z}_1\bar{Z}_2\}$$
$$+E\{Z_1\bar{Z}_2\}E\{Z_2\bar{Z}_1\})$$

而

$$\frac{1}{N}E|Z_1|^2=\frac{1}{N}E\left|\sum_{k=1}^{N}X(k)e^{-ik\lambda_0}\right|^2$$

$$=\int_{-\pi}^{\pi}\frac{\sin^2\frac{N}{2}(\lambda-\lambda_0)}{N\sin^2\frac{1}{2}(\lambda-\lambda_0)}f(\lambda)d\lambda$$

$$\leqslant 2\pi\max_{\lambda} f(\lambda)$$

同样可证

$$\frac{1}{N}E|Z_2|^2\leqslant 2\pi\max_{\lambda} f(\lambda)$$

因此

$$\frac{1}{N}|E\{Z_1Z_2\}|\leqslant\frac{1}{N}(E|Z_1|^2E|Z_2|^2)^{\frac{1}{2}}$$

$$\leqslant 2\pi\max_{\lambda} f(\lambda)$$

对上式中 Z_1, Z_2 用 $\bar{Z}_1, \bar{Z}_2$ 代替时亦有类似的估计，所以

$$E\left|\sum_{n=1}^{N-|\nu|}(\hat{m}(n)-m(n))X(n+|\nu|)\right|^2$$

$$\leqslant 3(2\pi\max_{\lambda} f(\lambda))^2$$

对(10)中的其它两项亦可建立同样的估计，所以存在与 N, ν 无关的常数 K，使(9)成立。定理证毕。

定理 1 将帮助我们建立 $\hat{B}_{\hat{X}}(\nu)$ 和 $\hat{B}(\nu)$ 有相同的渐近均值，方差和渐近分布，它表明当 $N\to\infty$ 时，

$$\sqrt{N}(\hat{B}_{\hat{X}}(\nu)-\hat{B}(\nu))\xrightarrow{P}0 \tag{11}$$

其中 $\xrightarrow{P}$ 表依概率收敛的极限，所以下面我们将只对 $\hat{B}(\nu)$ 来建立其渐近性质，或者相当于在以后的定理证明时都假定 $m(n)\equiv 0$. 而在实用中，当 $m(n)\neq 0$ 时，可用它的最小二乘方估计来代替，这时 $B(\nu)$ 估计的渐近性质是不变的。

定理 2　若 $\{X(n)\}$ 为实平稳正态过程，$\hat{B}(k)$ 由(5)规定，则

$$E\{\hat{B}(k)\}=\frac{N-|k|}{N}B(k) \tag{12}$$

若$\{X(n)\}$具有谱密度$f(\lambda)$，$\lambda\in(-\pi,\pi)$．且

$$\int_{-\pi}^{\pi} f^2(\lambda)d\lambda<\infty \tag{13}$$

或等价地

$$\sum_{k=-\infty}^{\infty} B^2(k)<\infty \tag{14}$$

则

$$\begin{aligned}
&\lim_{N\to\infty} N\mathrm{cov}(\hat{B}(k),\hat{B}(l))\\
&=\sum_{n=-\infty}^{\infty} B(n)B(n-k-l)+\sum_{n=-\infty}^{\infty} B(n)B(n+k+l)\\
&=4\pi\int_{-\pi}^{\pi}\cos l\lambda\cdot\cos k\lambda\cdot f^2(\lambda)d\lambda
\end{aligned} \tag{15}$$

$$\begin{aligned}
&\lim_{N\to\infty} ND\{\hat{B}(k)\}\\
&=\sum_{n=-\infty}^{\infty}(B^2(n)+B(n)B(n+2k))\\
&=4\pi\int_{-\pi}^{\pi}\cos^2 k\lambda f^2(\lambda)d\lambda
\end{aligned} \tag{16}$$

[证明]　利用$\hat{B}(k)$的表示式(5)，可得

$$\begin{aligned}
E\{\hat{B}(k)\}&=E\left\{\frac{1}{N}\sum_{n=1}^{N-|k|}X(n)X(n+|k|)\right\}\\
&=\frac{1}{N}\sum_{n=1}^{N-|k|}E\{X(n)X(n+|k|)\}\\
&=\frac{N-|k|}{N}B(k)
\end{aligned}$$

即(12)成立．若$X(n)$有谱密度$f(\lambda)$，则由协方差函数的谱分解

$$B(k)=\int_{-\pi}^{\pi} e^{ik\lambda}f(\lambda)d\lambda$$

利用富利埃级数中的巴塞伐尔(Paserval) 等式及$X(n)$为实过程

可得

$$\sum_{k=-\infty}^{\infty} B^2(k) = \sum_{k=-\infty}^{\infty} |B(k)|^2$$

$$= 2\pi \int_{-\pi}^{\pi} f^2(\lambda) d\lambda$$

所以(13),(14)是等价的. 由于 $B(-k)=B(k)$, $\hat{B}(-k)=\hat{B}(k)$, 不妨考虑 $k, l>0$ 的情形,利用(2)可得

$$E\{\hat{B}(k)\hat{B}(l)\}$$

$$= \frac{1}{N^2} E\left\{\left(\sum_{m=1}^{N-k} X(m)X(m+k)\right)\left(\sum_{n=1}^{N-l} X(n)X(n+l)\right)\right\}$$

$$= \frac{1}{N^2} \sum_{m=1}^{N-k} \sum_{n=1}^{N-l} E\{X(m)X(m+k)X(n)X(n+l)\}$$

$$= \frac{1}{N^2} \sum_{m=1}^{N-k} \sum_{n=1}^{N-l} (B(k)B(l) + B(m-n)B(m-n+k-l) + B(m-n-l)B(m+k-n))$$

$$N\operatorname{cov}(\hat{B}(k)\hat{B}(l))$$

$$= \frac{1}{N} \sum_{m=1}^{N-k} \sum_{n=1}^{N-l} (B(m-n)B(m-n+k-l) + B(m-n-l)B(m+k-n))$$

$$= \frac{1}{N} \sum_{n=1}^{N-l} \sum_{\nu=1-n}^{N-k-n} (B(\nu)B(\nu+k-l) + B(\nu-l)B(\nu+k))$$

$$= \frac{1}{N} \sum_{\nu=-(N-l-1)}^{N-k-1} \left(\sum_{n=1}^{\min(N-k-\nu, N-l)} 1\right) \cdot (B(\nu)B(\nu+k-l) + B(\nu-l)B(\nu+k))$$

由于(14)

$$\sum_{\nu=-\infty}^{\infty} |B(\nu)B(\nu+k-l)| < \infty$$

$$\sum_{\nu=-\infty}^{\infty} |B(\nu-l)B(\nu+k)| < \infty$$

又因

$$\left|\frac{1}{N}\sum_{n=1}^{\min(N-k-\nu, N-l)} 1\right| \leqslant 1$$

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{\min(N-k-\nu, N-l)} 1 = 1$$

所以

$$\lim_{N\to\infty} N\operatorname{cov}(\hat{B}(k), \hat{B}(l))$$

$$= \sum_{\nu=-\infty}^{\infty} (B(\nu)B(\nu+k-l) + B(\nu-l)B(\nu+k))$$

$$= \sum_{n=-\infty}^{\infty} (B(n)B(n+k-l) + B(n)B(n+k+l))$$

又利用巴塞伐尔等式及 $X(n)$ 为实过程，$f(-\lambda)=f(\lambda)$ 这一事实，可得

$$\sum_{n=-\infty}^{\infty} (B(n)B(n+k-l) + B(n)B(n+k+l))$$

$$= 2\pi\int_{-\pi}^{\pi} (e^{-i(k-l)\lambda} + e^{-i(k+l)\lambda}) f^2(\lambda)d\lambda$$

$$= 2\pi\int_{-\pi}^{\pi} (\cos(k-l)\lambda + \cos(k+l)\lambda) f^2(\lambda)d\lambda$$

$$= 4\pi\int_{-\pi}^{\pi} \cos k\lambda\cos l\lambda f^2(\lambda)d\lambda$$

即(15)成立，令 $k=l$ 时，即可由(15)推出(16)。

注．若对实过程 $X(n)$，考虑到其谱密度满足 $f(-\lambda)=f(\lambda)$，取 $f_1(\lambda)=2f(\lambda)$，则

$$B(n) = \int_0^{\pi} \cos n\lambda f_1(\lambda)d\lambda$$

此时(15)，(16)可写为

$$\lim_{N\to\infty} N\operatorname{cov}(\hat{B}(k), \hat{B}(l))$$

$$= 2\pi \int_0^\pi \cos k\lambda \cos l\lambda f_1^2(\lambda) d\lambda \tag{17}$$

$$\lim_{N\to\infty} ND\{\hat{B}(k)\}$$

$$= 2\pi \int_0^\pi (1+\cos 2k\lambda) f_1^2(\lambda) d\lambda \tag{18}$$

系　对于$\hat{B}_{\hat{X}}(k)$也成立类似的等式，即

$$\lim_{N\to\infty} \sqrt{N} |E\{\hat{B}_{\hat{X}}(\nu)\} - B(\nu)| = 0 \tag{19}$$

$$\lim_{N\to\infty} N\operatorname{cov}(\hat{B}_{\hat{X}}(k), \hat{B}_X(l))$$

$$= \sum_{n=-\infty}^{\infty} (B(n)B(n+k-l) + B(n)B(n+k+l))$$

$$= 4\pi \int_{-\pi}^{\pi} \cos k\lambda \cos l\lambda f^2(\lambda) d\lambda \tag{20}$$

$$\lim_{N\to\infty} ND\{\hat{B}_{\hat{X}}(k)\}$$

$$= \sum_{n=-\infty}^{\infty} (B(n) + B(n)B(n+2k))$$

$$= 4\pi \int_{-\pi}^{\pi} \cos^2 k\lambda f^2(\lambda) d\lambda \tag{21}$$

［证明］　因为

$$E|\hat{B}_{\hat{X}}(k) - \hat{B}(k)|^2$$

$$= (E\{\hat{B}_{\hat{X}}(k)\} - E\{\hat{B}(k)\})^2 + D\{\hat{B}_{\hat{X}}(k) - \hat{B}(k)\}$$

所以由定理 1

$$\lim_{N\to\infty} ND\{\hat{B}_{\hat{X}}(k) - B(k)\} = 0 \tag{22}$$

$$\lim_{N\to\infty} N(E\{\hat{B}_X(k)\} - E\{\hat{B}(k)\})^2 = 0$$

由(12)及上式即得(19)．若对随机变量η，记

$$\|\eta\|^2=D\{\eta\}$$

则

$$|D\{\hat{B}_{\hat{X}}(k)\}-D\{\hat{B}(k)\}|=|\|\hat{B}_{\hat{X}}(k)\|^2-\|\hat{B}(k)\|^2|$$

$$\leqslant\|\hat{B}_{\hat{X}}(k)-\hat{B}(k)\|\cdot(2\|\hat{B}(k)\|+\|\hat{B}_{\hat{X}}(k)-\hat{B}(k)\|)$$

由(16)，$\sqrt{N}\|\hat{B}(k)\|$是有界的，所以由上式及(22)可得

$$\lim_{N\to\infty}N[D\{\hat{B}_{\hat{X}}(k)\}-D\{\hat{B}(k)\}]=\lim_{N\to\infty}N(\|\hat{B}_{\hat{X}}(k)\|^2-\|\hat{B}(k)\|^2)=0$$

由上式及(16)即推出(21)成立，类似于上面的可证明

$$\lim_{N\to\infty}N|\operatorname{cov}(\hat{B}_{\hat{X}}(k),\hat{B}_{\hat{X}}(l))-\operatorname{cov}(\hat{B}(k),\hat{B}(l))|=0$$

这样由(15)亦可推出(19)成立.

从定理 2 可以看出，$\hat{B}(k)$不是 $B(k)$的无偏估计，$\hat{B}^*(k)$才是 $B(k)$的无偏估计，尽管如此，在时间序列分析中还是较多地采用 $\hat{B}(k)$或 $\hat{B}_{\hat{X}}(k)$作为 $B(k)$的估计．由定理 2 不难推出．

系　$\hat{B}(k)$，$\hat{B}_{\hat{X}}(k)$都是 $B(k)$的一致性估计，且

$$\lim_{N\to\infty}E|\hat{B}(k)-B(k)|^2=\lim_{N\to\infty}E|\hat{B}_{\hat{X}}(k)-B(k)|^2=0$$

因为由(5)规定的 $\hat{B}(k)$是N维正态变量 $X(1),\cdots,X(N)$的二次型，所以要讨论 $\hat{B}(k)$的渐近分布之前，我们先来介绍关于正态变量二次型分布的下列引理.

引理 1　若N维正态变量 $X=(X(1),\cdots,X(N))'$ 为$N(0,B)$分布，又 A 为 $N\times N$ 对称阵，则存在相互独立 $N(0,1)$ 分布的正态随机变量 $\xi(1),\cdots,\xi(N)$，使X的二次型 $Y_N=X'AX$ 表为

$$Y_N=X'AX=\sum_{j=1}^{N}\lambda_j^{(N)}\xi^2(j)$$

其中 $\lambda_1^{(N)},\cdots,\lambda_j^{(N)}$ 为对称阵 $B^{\frac{1}{2}}AB^{\frac{1}{2}}$ 的全部特征根（按重数重复计算）.

［证明］ 首先由多维正态变量的性质，对 X 必存在 $N(0,1)$ 分布的 N 维正态变量 η，使 $X=B^{\frac{1}{2}}\eta$，这时 $Y_N=\eta' B^{\frac{1}{2}}AB^{\frac{1}{2}}\eta$，而由线代数中二次型化标准型的结果可知必存在正交阵 U，使

$$U'B^{\frac{1}{2}}AB^{\frac{1}{2}}U=\begin{bmatrix}\lambda_1 & & 0\\ & \ddots & \\ 0 & & \lambda_N\end{bmatrix} \tag{23}$$

这时，若取

$$U'\eta=\xi=\begin{bmatrix}\xi(1)\\ \vdots\\ \xi(N)\end{bmatrix}$$

则

$$\eta=U\xi$$

$$Y_N=\eta' B^{\frac{1}{2}}AB\eta=\xi'U' B^{\frac{1}{2}}ABU\xi$$

$$=\sum_{j=1}^{N}\lambda_j^{(N)}\xi^2(j)$$

而由(23)可知，$\lambda_1^{(N)},\cdots,\lambda_N^{(N)}$ 为 $B^{\frac{1}{2}}AB^{\frac{1}{2}}$ 的全部特征根.

引理 2 对引理 1 中的 X,Y_N，若满足

$$\lim_{N\to\infty}\max_{1\leqslant j\leqslant N}|\lambda_j^{(N)}|/D\{Y_N\}=0 \tag{24}$$

则 Y_N 为渐近正态分布的，即

$$\lim_{N\to\infty}P\left(\frac{Y_N-E\{Y_N\}}{\sqrt{D\{Y_N\}}}<x\right)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-\frac{u^2}{2}}du \tag{25}$$

［证明］ 由引理 1

$$\zeta_N=\frac{Y_N-E\{Y_N\}}{\sqrt{D\{Y_N\}}}$$

$$=\sum_{j=1}^{N}\frac{\lambda_j^{(N)}}{\sqrt{D\{Y_N\}}}(\xi^2(j)-1)$$

$$=\sum_{j=1}^{N}\xi_{Nj}$$

由于 $\xi^2(j)-1$ 是均值为零，且相互独立同分布的，再由(3)，$D(Y_N)=2\sum_{j=1}^{N}(\lambda_j^{(N)})^2$. 我们来验证 $\zeta_N=\sum_{j=1}^{N}\xi_{Nj}$ 是满足林德伯格条件的，

$$\sum_{j=1}^{N}\int_{|\xi_{Nj}|\geq e}\xi^2_{Nj}dP$$

$$=\sum_{j=1}^{N}\frac{(\lambda_j^{(N)})^2}{D\{Y_N\}}\int_{|\xi^2(j)-1|\geq\frac{\sqrt{D\{Y_N\}}}{|\lambda_j^{(N)}|}e}(\xi^2(j)-1)^2dP$$

$$\leqslant\frac{1}{2}\int_{|\xi^2(1)-1|\geq\frac{\sqrt{D\{Y_N\}}}{\max\limits_{1\leq j\leq N}|\lambda_j^{(N)}|}e}(\xi^2(1)-1)^2dP$$

所以由(24)，对任一 $e>0$

$$\lim_{N\to\infty}\sum_{j=1}^{N}\int_{|\xi_{Nj}|\geq e}\xi^2_{Nj}dP=0$$

因此，由中心极限定理可知(25)成立.

为了估计 $\lambda_j^{(N)}$，我们介绍下列结果.

引理 3 若 $N\times N$ 厄密特阵 $C^{(N)}=(C_{k-l})$，而

$$C_{k-l}=\int_{-\pi}^{\pi}e^{i(k-l)\lambda}g(\lambda)d\lambda,\quad 1\leqslant k,l\leqslant N$$

又以 $\lambda(C^{(N)})$ 表示 $C^{(N)}$ 特征根中模最大者，则当 $g(\lambda)$ 有界时

$$\lambda(C^{(N)})\leqslant 2\pi\sup_{\lambda}|g(\lambda)|$$

而当 $g(\lambda)$ 为 p 次可积时，

$$\lim_{N\to\infty}N^{-\frac{1}{p}}\lambda(C^{(N)})=0$$

[证明] 若记$\|x\|^2=\sum_{k=1}^{N}|x_k|^2$，则从厄密特阵特征根的性质可知

$$\lambda(C^{(N)})=\sup_{\|x\|=1}\left|\sum_{k,l=1}^{N}x_k\bar{x}_l C_{k-l}\right|$$

$$=\sup_{\|x\|=1}\left|\int_{-\pi}^{\pi}\left|\sum_{k=1}^{N}x_k e^{ik\lambda}\right|^2 g(\lambda)d\lambda\right|$$

$$\leqslant\sup_{\lambda}|g(\lambda)|\sup_{\|x\|=1}\int_{-\pi}^{\pi}\left|\sum_{k=1}^{N}x_k e^{ik\lambda}\right|^2 d\lambda$$

$$=2\pi\sup_{\lambda}|g(\lambda)|$$

当$g(\lambda)$为p次可积时，利用许伐兹不等式有

$$\left|\sum_{k=1}^{N}x_k e^{ik\lambda}\right|\leqslant\sum_{k=1}^{N}|x_k|^2\sum_{k=1}^{N}|e^{ik\lambda}|^2$$

$$\leqslant N\|x\|$$

而

$$\lambda(C^{(N)})=\sup_{\|x\|=1}\left|\int_{-\pi}^{\pi}\left|\sum_{k=1}^{N}x_k e^{ik\lambda}\right|^2 g(\lambda)d\lambda\right|$$

$$\leqslant N\int_{|g|>\varepsilon N^{1/p}}|g(\lambda)|d\lambda+\varepsilon N^{1/p}$$

$$\cdot\left(\sup_{\|x\|=1}\int_{|g|\leqslant\varepsilon N^{1/p}}\left|\sum_{k=1}^{N}x_k e^{ik\lambda}\right|^2 d\lambda\right)$$

$$\leqslant\frac{N}{\varepsilon^{p-1}N^{1-\frac{1}{p}}}\int_{|g|>\varepsilon N^{1/p}}|g(\lambda)|^p d\lambda+\varepsilon N^{1/p}$$

$$=\frac{N^{1/p}}{\varepsilon^{p-1}}\int_{|g|>\varepsilon N^{1/p}}|g(\lambda)|^p d\lambda+\varepsilon N^{1/p}$$

所以

$$\overline{\lim_{N\to\infty}}N^{-\frac{1}{p}}\lambda(C^{(N)})\leqslant\varepsilon$$

由ε的任意性，可得引理成立。

有了上面的准备后，就可以推出关于 $\hat{B}(k)$ 渐近分布的下列定理.

定理 3 若$\{X(n)\}$为满足(13)或(14)的实平稳正态过程，$\hat{B}(k)$由(5)规定，则对任一 k_0，当N趋于无穷时，k_0 维随机变量

$$\sqrt{N}(\hat{B}(0)-B(0)),\ \sqrt{N}(\hat{B}(1)-B(1)),\\ \cdots,\sqrt{N}(\hat{B}(k_0-1)-B(k_0-1)) \tag{26}$$

的联合分布收敛于k_0维正态分布 $N(0,\Sigma)$，其中

$$\Sigma=(\sigma_{kl})$$

$$\sigma_{kl}=\sum_{n=-\infty}^{\infty}(B(n)B(n+k-l)+B(n)B(n+k+l))$$

$$=4\pi\int_{-\pi}^{\pi}\cos k\lambda\cos l\lambda f^2(\lambda)d\lambda,\quad 0\leqslant k,\ l\leqslant k_0-1 \tag{27}$$

［证明］ 为简单起见，我们只证明$\sqrt{N}(\hat{B}(k)-B(k))$的极限分布. 对(26)中的多维随机变量的极限分布问题要用到多维的中心极限定理，所以我们省略了. 因为，当 $k>0$ 时，令

$$Y_N\triangleq\sqrt{N}\hat{B}(k)$$

$$=\frac{1}{\sqrt{N}}\sum_{n=1}^{N-k}X(n)X(n+k)$$

$$=X^{\tau}A_NX \tag{28}$$

其中

$$A_N=(a_{j-l}^{(N)})$$

$$a_{j-l}^{(N)}=\begin{cases}\dfrac{1}{2\sqrt{N}}, & |j-l|=k\\ 0\quad, & |j-l|\neq k\end{cases}$$

所以取

$$g(\lambda)=\frac{1}{2\pi\sqrt{N}}\cos k\lambda \tag{29}$$

则有

$$a_{j-l}^{(N)}=\int_{-\pi}^{\pi}e^{i(j-l)\lambda}g(\lambda)\,d\lambda$$

为了对(28)中的 $X'A_NX$ 验证它满足引理2的条件，首先应注意到 X 的协方差阵 $B_N=(B(j-l))$，而

$$B(j-l)=\int_{-\pi}^{\pi}e^{i(j-l)\lambda}f(\lambda)\,d\lambda$$

若记 $x=(x_1,x_2,\cdots,x_N)'$，则由引理3及(29)

$$\lambda(B_N^{\frac{1}{2}}A_NB_N^{\frac{1}{2}})=\sup_{\sum_{j=1}^{N}x_j^2=1}x'B_N^{\frac{1}{2}}A_NB_N^{\frac{1}{2}}x$$

$$\leqslant\lambda(A_N)\sup_{\sum_{j=1}^{N}x_j^2=1}x'B_N^{\frac{1}{2}}B_N^{\frac{1}{2}}x$$

$$\leqslant\lambda(A_N)\lambda(B_N)$$

由引理3及(29)

$$\lambda(A_N)\leqslant\frac{1}{\sqrt{N}}$$

又由 $f(\lambda)$ 平方可积及引理3,

$$\lim_{N\to\infty}N^{-\frac{1}{2}}\lambda(B_N)=0$$

所以

$$\lambda(B_N^{\frac{1}{2}}A_NB_N^{\frac{1}{2}})\leqslant\lambda(A_N)\lambda(B_N)$$

$$\leqslant\frac{1}{\sqrt{N}}\lambda(B_N)$$

$$\lim_{N\to\infty}\lambda(B_N^{\frac{1}{2}}A_NB_N^{\frac{1}{2}})=0$$

另一方面，由定理2

$$\lim_{N\to\infty}D\{Y_N\}=\lim_{N\to\infty}D\{\sqrt{N}\hat{B}(k)\}$$

$$=\sigma_{kk}>0$$

所以对 $Y_N=\sqrt{N}\hat{B}(k)$, (24)成立, 即

$$\lim_{N\to\infty}P\left\{\frac{\sqrt{N}\hat{B}(k)-E\{\sqrt{N}\hat{B}(k)\}}{D\{\sqrt{N}\hat{B}(k)\}}<x\right\}$$
$$=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-\frac{u^2}{2}}du$$

又由定理 2, (12)及(16)关于 $E\{\hat{B}(k)\}$ 和 $D\{\hat{B}(k)\}$ 的渐近表示式, 不难推出

$$\lim_{N\to\infty}P\left\{\frac{\sqrt{N}(\hat{B}(k)-B(k))}{\sqrt{\sigma_{kk}}}<x\right\}$$
$$=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-\frac{u^2}{2}}du \tag{30}$$

这样就得到了 $\hat{B}(k)$ 的渐近分布. 定理证毕.

利用定理 1 和定理 2 的结论, 不难推出对 $\hat{B}_{\hat{X}}(k)$ 定理 3 的结论同样是成立的, 所以有

系 若 $\{X(n)\}$ 满足定理 3 的条件, $\hat{B}_{\hat{X}}(k)$ 由(8) 规定, 则对任一 k_0 当 N 趋于无穷时, k_0 维随机变量

$$\sqrt{N}\ (\hat{B}_{\hat{X}}(0)-B(0)),\ \sqrt{N}\ (\hat{B}_{\hat{X}}(1)-B(1)),$$
$$\cdots, \sqrt{N}\ (\hat{B}_{\hat{X}}(k_0-1)-B(k_0-1))$$

的联合分布收敛于 k_0 维正态分布 $N(0,\ \Sigma)$, 其中 $\Sigma=(\sigma_{kl})$, 而 σ_{kl} 由(20)规定.

上面我们只是对协方差函数估计量讨论了它的渐近性, 当我们涉及到几个实平稳序列时, 若 $\{X(n)\}$, $\{Y(n)\}$ 是均值为零且平稳相关的, 即对一切 n 有

$$E\{X(n)\}=E\{Y(n)\}=0$$
$$E\{X(n+k)Y(n)\}=B_{XY}(k)$$

$B_{XY}(n)$ 又称 X, Y 的互协方差函数. 对于它, 亦可类似于(5)取

$$\hat{B}_{XY}(k)=\frac{1}{N}\sum_{n=1}^{N-|k|}X(n+k)Y(n) \tag{31}$$

作为 $B_{XY}(k)$ 的估计量，对于互协方差函数，亦可用上面相仿的办法得到类似于定理 2 和 3 的结果．例如有

定理 4　若 $\{X(n)\}$, $\{Y(n)\}$ 是均值为零且平稳相关的实联合正态平稳序列，且分别存在平方可积的谱密度函数 $f_{XX}(\lambda)$ 和 $f_{YY}(\lambda)$．$\hat{B}_{XY}(k)$ 由(33)规定，则对任一 k_0，当 N 趋于无穷时，k_0 维随机变量

$$\sqrt{N}(\hat{B}_{XY}(0)-B_{XY}(0)),\ \sqrt{N}(\hat{B}_{XY}(1)-B_{XY}(1)),$$
$$\cdots,\ \sqrt{N}(\hat{B}_{XY}(k_0-1)-B_{XY}(k_0-1))$$

的联合分布收敛于 k_0 维正态分布 $N(0,\Sigma)$，其中

$$\Sigma=(\sigma_{kl})$$

$$\sigma_{kl}=\sum_{n=-\infty}^{\infty}[B_{XX}(n)B_{YY}(n+k-l)+B_{XY}(n)B_{XY}(-n+k+l)]$$
$$=2\pi\int_{-\pi}^{\pi}(\cos(k-l)\lambda f_{XX}(\lambda)f_{YY}(\lambda)+\cos(k+l)\lambda|f_{XY}(\lambda)|^2)d\lambda,$$
$$0\leqslant k,\ l\leqslant k_0-1 \tag{32}$$

而 $f_{XY}(\lambda)$ 为 X,Y 的交互谱密度．

二、功率谱密度的估计

我们已经知道，对于均值为零的平稳时间序列来说，它的谱函数和协方差函数一样，完整地描述了它的二阶统计特征，虽然由协方差函数也可唯一地确定对应的谱函数，但由于谱函数的特殊作用，我们还需要专门讨论它的估计问题．一方面由于谱的物理含义，例如对描述随机振动的平稳过程来说，我们求得它的谱以后就可看出各种不同频率的振动其功率的分布．这使我们有需要求得随机过程的谱，另一方面，由(4.3.27)，我们可以通过线性系统输入输出的谱和交互谱来求出系统的频率特征，为了做到这一点也必须先按输入输出资料对谱和交互谱进行估计，所以我们有必要

对随机过程谱的估计问题进行专门的讨论. 在平稳过程谱估计中, 以谱密度的估计使用较多, 我们的讨论也将集中在这方面.

关于谱密度的估计, 通常可有两种模型, 一种是有限参数模型, 即假定待估计的谱密度其函数形式已知而只含有限个未知参数, 这时通过对这有限个参数的估计可求得谱密度的估计, 这种做法将在下一节进行介绍. 另一种是非参数模型, 即并不预先假定谱密度的函数形式, 而用协方差序列或其它子样的二次型来对谱密度进行估计. 近年来又由于快速富利埃算法的出现, 在实现上又有从子样协方差序列的富利埃变换或用子样序列富利埃变换加窗等不同的做法, 这方面的基本统计考虑我们将在这一节介绍.

在第四章, 我们已指出了平稳过程协方差函数和谱密度间的相互关系. 特别当其协方差函数 $B(n)$ 满足 $\sum\limits_{n=-\infty}^{\infty}|B(n)|<\infty$ 时, 则必存在谱密度 $f(\lambda)$, 且有

$$B(n)=\int_{-\pi}^{\pi}e^{in\lambda}f(\lambda)d\lambda$$

$$f(\lambda)=\frac{1}{2\pi}\sum_{n=-\infty}^{\infty}B(n)e^{-in\lambda} \tag{33}$$

在上面, 我们已经讨论过 $B(n)$ 的估计量 $\hat{B}(n)$, 所以在这里自然就想到用(33)作为我们估计 $f(\lambda)$ 的出发点, 若我们从实 X_T 出发来估计它的谱密度 $f(\lambda)$, 当获得观测 $\{X(1), X(2), \cdots, X(N)\}$ 之后, 就可作出

$$\hat{B}(k)=\frac{1}{N}\sum_{n=1}^{N-|k|}X(n)X(n+|k|),\ |k|\leqslant N-1$$

所以按(33)作为 $f(\lambda)$ 的估计量就可以取

$$I_N(\lambda)=\frac{1}{2\pi}\sum_{k=-N+1}^{N-1}\hat{B}(k)e^{-ik\lambda} \tag{34}$$

若将 $\hat{B}(k)$ 的表示式代入(34)并认为

$$X(n)=0,\ \text{当}\quad n\leqslant 0\quad\text{或}\quad n\geqslant N+1 \tag{35}$$

则有

$$\begin{aligned}
I_N(\lambda)&=\frac{1}{2\pi N}\sum_{k=-N+1}^{N-1}\left(\sum_{n=1}^{N-|k|}X(n)X(n+|k|)\right)e^{-ik\lambda}\\
&=\frac{1}{2\pi N}\sum_{k=-\infty}^{\infty}\sum_{n=-\infty}^{\infty}X(n)X(n+k)e^{-ik\lambda}\\
&=\frac{1}{2\pi N}\left|\sum_{k=-\infty}^{\infty}X(n)\,e^{-in\lambda}\right|^2\\
&=\frac{1}{2\pi N}\left|\sum_{k=1}^{N}X(n)\,e^{-in\lambda}\right|^2
\end{aligned} \tag{36}$$

由(2.30)它就是 $X(n)$ 的周期图．这是一个直观上容易得到的 X_T 谱密度 $f(\lambda)$ 的估计量，但进一步的统计分析将表明它是并不满意的，为了说明这一点，我们对 X_T 为白噪声来计算它的均值和方差，因为当 X_T 为白噪声时，

$$f(\lambda)=\frac{\sigma^2}{2\pi}$$

$$B(k)=\sigma^2\delta_{0k}$$

这时$\{X(1),\cdots,X(N)\}$是相互独立 $N(0,\sigma^2)$分布的，所以由直接计算不难得出

$$\begin{aligned}
\eta_j&=\frac{1}{\sqrt{N}}\sum_{k=1}^{N}X(k)e^{-i2\pi k\frac{j}{N}}\\
&=\alpha_j+i\beta_j,\quad 1\leqslant j\leqslant\frac{N}{2}
\end{aligned} \tag{37}$$

是相互独立的，且 α_j，β_j 也是相互独立的 $N\left(0,\dfrac{\sigma^2}{2}\right)$分布，

$$\begin{aligned}
I_N\left(\frac{j}{N}2\pi\right)&=\frac{1}{2\pi}|\eta_j|^2\\
&=\frac{1}{2\pi}(\alpha_j^2+\beta_j^2)
\end{aligned}$$

所以

$$E\left\{I_N\left(\frac{j}{N}2\pi\right)\right\}=\frac{\sigma^2}{2\pi}=f(\lambda) \tag{38}$$

$$D\left\{I_N\left(\frac{j}{N}2\pi\right)\right\}=\frac{1}{4\pi^2}[E\{|\eta_j|^4\}-(E\{|\eta_j|^2\})^2]$$

$$=\frac{\sigma^4}{4\pi^2} \tag{39}$$

(38)表明 $I_N(\lambda)$ 作为 $f(\lambda)=\frac{\sigma^2}{2\pi}$ 的估计是无偏的，而(39)表明这个估计不是(均方)一致的．当 X_T 为一般的平稳正态过程时，若其有连续谱密度为 $f(\lambda)$，则也可进行类似的计算得到如下的结果：

$$\begin{aligned}&\lim_{N\to\infty}E\{I_N(\lambda)\}=f(\lambda)\\&\lim_{N\to\infty}D\{I_N(\lambda)\}=f^2(\lambda)\\&\lim_{N\to\infty}\operatorname{cov}(I_N(\lambda),I_N(\mu))=0,\quad\lambda\neq\mu\end{aligned} \tag{40}$$

即用周期图来估计谱时是渐近无偏的，但不是均方一致的，随 N 无限增大，估计量的方差却不会逐渐减小到零．对这一点，在实用上是不满意的，所以要进一步寻求更好的估计量．

当 X_T 为白噪声时，由上面的讨论已经知道，$I_N\left(\frac{j}{N}2\pi\right)$ 是互不相关的，所以当 $\lambda_0=\frac{j_0}{N}$ 时，作

$$\tilde{I}\left(\frac{j_0}{N}2\pi\right)=\frac{1}{2L+1}\sum_{j=-L}^{L}I\left(\frac{j_0+j}{N}2\pi\right) \tag{41}$$

那么

$$E\left\{\tilde{I}\left(\frac{j_0}{N}2\pi\right)\right\}=\frac{\sigma^2}{2\pi}=f(\lambda)$$

$$D\left\{\tilde{I}\left(\frac{j_0}{N}2\pi\right)\right\}=\frac{1}{(2L+1)^2}\sum_{j=-L}^{L}D\left\{I\left(\frac{j_0+j}{N}2\pi\right)\right\}$$

$$=\frac{1}{2L+1}\cdot\frac{\sigma^4}{4\pi^2}$$

这样，为了估计 $f(\lambda_0)$，若利用 $I'_N(\lambda)$ 在 λ 邻近几点的值的平均来作为估计量时，无偏性不会改变，而估计量的方差就会减小，而且被平均的点越多，方差减小得越多．当 N 无限增大时，是有可能多取几点来平均的．对于一般的平稳过程，因为(40)保证了对不同的 λ，周期图 $I_N(\lambda)$ 估计量渐近地互不相关，所以也可以用类似的想法来消除 $I_N(\lambda)$ 的不一致性，并构造 $f(\lambda)$ 的估计量．为此我们可以考虑下列形式的估计量

$$f_N(\lambda)=\int_{-\pi}^{\pi} W_N(\lambda-\mu)I_N(\mu)d\mu \tag{42}$$

上述积分就是对 $I_N(\mu)$ 的平均，而其中 $W_N(\mu)$ 就起一个加权的作用，它又称为频率域上的**谱窗函数**．在估计 $f(\lambda)$ 时，它就按 (42) 把 $f(\mu)$ 在 λ 的某个邻域内加权平均．当然平均的效果一方面是比 $I_N(\lambda)$ 减小了方差．另一方面把异于 λ 点上的 $I_N(\mu)$ 也取来平均，这样就增大了估计量的偏差 $E\{\hat{f}(\lambda)\}-f(\lambda)$，所以从减小估计量方差来说，$W_N(\mu)$ 能在较大的范围内取平均更好，而从减小偏差来看，$W_N(\mu)$ 取平均的范围又不能太大，因此还必须兼顾这两方面的要求，若 $W_N(\mu)$ 取下列较简单的形式

$$W_N(\mu)=\begin{cases}\dfrac{1}{2B_N}, & |\mu|<B_N\\ 0, & |\mu|\geqslant B_N\end{cases}$$

其中 B_N 表示取平均范围的宽度，一般地，随 N 趋于无穷时 B_N 应趋于零．这时

$$\begin{aligned}\hat{f}_N(\lambda)&=\frac{1}{2B_N}\int_{-B_N}^{B_N} I_N(\lambda+\mu)d\mu\\&=\frac{1}{4B_N\pi}\int_{-B_N}^{B_N}\sum_{k=-N+1}^{N-1}\hat{B}(k)e^{-ik(\lambda+\mu)}d\mu\\&=\frac{1}{2\pi}\sum_{k=-N+1}^{N-1}\hat{B}(k)\frac{\sin kB_N}{kB_N}e^{-ik\lambda}\end{aligned} \tag{43}$$

所以对$I_N(\lambda)$平均的效果，从时间域上来看，就相当于将$\hat{B}(k)$乘上一个因子w_{Nk}:

$$w_{Nk}=\frac{\sin kB_N}{kB_N}$$

这表明以$\hat{B}(k)w_{Nk}$按(43)来计算$f(\lambda)$的估计量$\hat{f}(\lambda)$也有可能减小估计量的方差，解决周期图估计$I_N(\lambda)$的不一致性问题. 虽然(43)和(42)是相同的，但若我们从$\hat{B}(k)$出发计算$\hat{f}_N(\lambda)$时(43)更直接，所以下面我们就讨论如下形式的估计量

$$\hat{f}_N(\lambda)=\frac{1}{2\pi}\sum_{k=-N+1}^{N-1}\hat{B}(k)\,w_{Nk}e^{-ik\lambda} \tag{44}$$

在此，w_{Nk}又称为时间域上的**谱窗因子**，所以一种谱密度估计量就与一种谱窗因子相对应，常用的w_{Nk}往往可表为：

$$w_{Nk}=h(B_Nk) \tag{45}$$

在这一情况下频率域的谱窗函数为

$$W_N(\lambda)=\frac{1}{2\pi}\sum_{k=-\infty}^{\infty}w_{Nk}e^{-ik\lambda} \tag{46}$$

$h(x)$往往满足

$$h(x)=0,\quad |x|\geqslant a$$

所以(46)中的和式往往只包含有限项，今后我们将讨论这一类谱密度估计的渐近性质，并且假定$h(x)$是$(-\infty,\ \infty)$上连续有界的可积函数，它对某个$q>0$，还存在下列极限

$$\lim_{x\to 0}\frac{1-h(x)}{|x|^q}=h_q,\quad |h(x)|\leqslant M \tag{47}$$

且

$$\sum_{n=-\infty}^{\infty}|n|^q|B(n)|<\infty \tag{48}$$

则有

定理5 若零均值实平稳正态过程 $X_T=\{X(n),\quad n=0,$

$\pm 1, \cdots\}$有连续谱密度 $f(\lambda)$,若由(44)规定的 $\hat{f}_N(\lambda)$ 为它的估计量,且若随 N 趋于无穷, B_N^{-1}, NB_N, NB_N^q 都趋于无穷,则当(45)——(48)成立时便有

$$\lim_{N\to\infty} B_N^{-q}(E\{\hat{f}_N(\lambda)\}-f(\lambda))$$
$$=-\frac{h_q}{2\pi}\sum_{n=-\infty}^{\infty}|n|^q B(n)e^{-in\lambda} \tag{49}$$

$$\lim_{N\to\infty} NB_N \mathrm{cov}(\hat{f}_N(\lambda_1), \hat{f}_N(\lambda_2))$$
$$=\begin{cases} 0, & \lambda_1 \neq \pm\lambda_2 \\ f^2(0)\int_{-\infty}^{\infty} h^2(x)dx, & \lambda_1=\pm\lambda_2\neq 0 \\ 2f^2(0)\int_{-\infty}^{\infty} h^2(x)dx, & \lambda_1=\lambda_2=0 \end{cases} \tag{50}$$

[证明] 由(44), (45)及(12),

$$E\{\hat{f}_N(\lambda)\}-f(\lambda)$$
$$=\frac{1}{2\pi}\sum_{k=-N+1}^{N-1}\left(1-\frac{|k|}{N}\right)B(k)h(B_N k)e^{-ik\lambda}$$
$$-\frac{1}{2\pi}\sum_{k=-\infty}^{\infty}B(k)e^{-ik\lambda}$$
$$=\frac{1}{2\pi}\sum_{|k|\leq N-1}(h(B_N k)-1)B(k)e^{-ik\lambda}$$
$$+\frac{1}{2\pi}\sum_{|k|\leq N-1}\frac{|k|}{N}B(k)h(B_N k)e^{-ik\lambda}$$
$$-\frac{1}{2\pi}\sum_{|k|\geq N}B(k)e^{-ik\lambda}$$
$$=I_1+I_2+I_3$$

而

$$B_N^{-q}I_1=\frac{1}{2\pi}\sum_{|k|\leq N-1}\frac{h(B_N k)-1}{|B_N k|^q}\cdot|k|^q B(k)e^{-ik\lambda}$$

由(47), (48),

$$\lim_{N\to\infty} B_N^{-q} I_1 = -\frac{h_q}{2\pi}\sum_{n=-\infty}^{\infty}|n|^q B(n)e^{-in\lambda}$$

又因

$$|B_N^{-q} I_2| \leqslant \frac{B_N^{-q} N^{-1} M}{2\pi}\sum_{|k|\leqslant N-1}|k|\,|B(k)|$$

$$\leqslant\begin{cases}(B_N N)^{-q}\dfrac{M}{2\pi}\sum\limits_k |k|^q|B(k)|, & q\leqslant 1\\ B_N^{-q} N^{-1}\cdot\dfrac{M}{2\pi}\sum\limits_k |k|^q|B(k)|, & q\geqslant 1\end{cases}$$

$$|B_N^{-q} I_3| \leqslant \frac{B_N^{-q}\cdot N^{-q}}{2\pi}\sum_{|k|\geqslant N}|k|^q|B(k)|$$

所以

$$\lim_{N\to\infty} B_N^{-q} I_2 = 0, \qquad \lim_{N\to\infty} B_N^{-q} I_3 = 0 \tag{51}$$

这样(49)是成立的.

至于(50)的证明, 可以运用定理 2 的方法直接计算 $\mathrm{cov}(\hat{f}_N(\lambda_1), \hat{f}_N(\lambda_2))$并估计当$N$趋于无穷时的极限, 但由于计算较为冗长, 所以这里省略了①. 同样可以证明, 当 $Z_T=\{Z(n)\}$满足(1)时, 则(44)中的 $\hat{B}(k)$代以 $\hat{B}_{\hat{X}}(k)$时定理 4 仍然是成立的.

下面我们就来讨论一些具体的谱窗因子和它们所对应的谱密度估计量.

[例 1] 若取

$$h(x)=\frac{\sin x}{x}, \qquad w_{Nk}=\frac{\sin B_N k}{B_N k}$$

则

$$W_N(\lambda)=\begin{cases}\dfrac{1}{2B_N}, & |\lambda|\leqslant B_N\\ 0, & |\lambda|>B_N\end{cases}$$

① 可参见 E. J. Hannen, Multiple Time Series 一书.

$$\hat{f}_N(\lambda)=\frac{1}{2\pi}\sum_{k=-N+1}^{N-1}\hat{B}(k)w_{Nk}e^{-ik\lambda}$$

$$=\int_{-\pi}^{\pi}W_N(\lambda-\theta)I_N(\theta)d\theta$$

$$=\frac{1}{2B_N}\int_{-B_N}^{+B_N}I_N(\lambda+\theta)d\theta \tag{52}$$

从(52)可看出，在这个估计量中，它是符合于对周期图进行平均这一直观想法的，而且由 $h(x)$ 的取法可知，当 $q=2$ 时，$h_2=\frac{1}{6}$. 因此对 $q\leqslant 2$ 时，只要 $\sum_{n=-\infty}^{\infty}|n|^q|B(n)|<\infty$，定理 4 的结论对它是适用的，但是，由于(52)中计算 $\hat{f}_N(\lambda)$ 时要用到全部 $\hat{B}(k)$，从计算上考虑这个窗是不常采用的.

［例 2］ 若取

$$h(x)=\begin{cases}1, & |x|\leqslant 1\\ 0, & |x|>1\end{cases}$$

$$w_{Nk}=\begin{cases}1, & |k|\leqslant\frac{1}{B_N}\\ 0, & |k|\geqslant\frac{1}{B_N}\end{cases}$$

则取 $M_N=\left[\frac{1}{B_N}\right]$ 时

$$W_N^{(2)}(\lambda)=\frac{\sin\frac{2M_n+1}{2}\lambda}{\sin\frac{\lambda}{2}} \tag{53}$$

$$\hat{f}_N(\lambda)=\frac{1}{2\pi}\sum_{k=-M_N}^{M_N}\hat{B}(k)e^{-ik\lambda}$$

$$=\int_{-\pi}^{\pi}\frac{\sin\frac{2M_n+1}{2}\theta}{\sin\frac{\theta}{2}}I_N(\lambda-\theta)d\theta \tag{54}$$

由于 $h(x)$ 的形式，这个估计量的窗又称**截断窗**，它对于任一 $q>0$，$h_q=0$，所以只要(48)成立，或者相当于 $f(\lambda)$ 本身足够光滑，$\hat{f}_N(\lambda)$ 估计 $f(\lambda)$ 的偏差可以渐近地很小，若 B_N 以很慢的速度趋于零，$D(\hat{f}_N(\lambda))$ 的方差也可与 $\frac{1}{\sqrt{N}}$ 相近，从计算方面来考虑，(54)表明它只需要用 $\hat{B}(k)$，$|k|\leqslant M_N=\left[\frac{1}{B_N}\right]$，不必计算全部 $\hat{B}(k)$，而且一般 $M_N\ll N$，所以在计算上也是可取的，但是它的频率窗函数 $W_N(\lambda)$ 如图 1 所示，除了在 $\lambda=0$ 附近有一个很陡的峰之外，边上还有不少正负相间的边瓣，这些边瓣在按(54)计算 $\hat{f}_N(\lambda)$ 时，就把 λ 附近的数值以较大的比例取来平均，而且还可能使 $\hat{f}_N(\lambda)$ 在某些点上变为负，这些都是它不足的地方，为了减进这方面的缺陷就有下列两种窗.

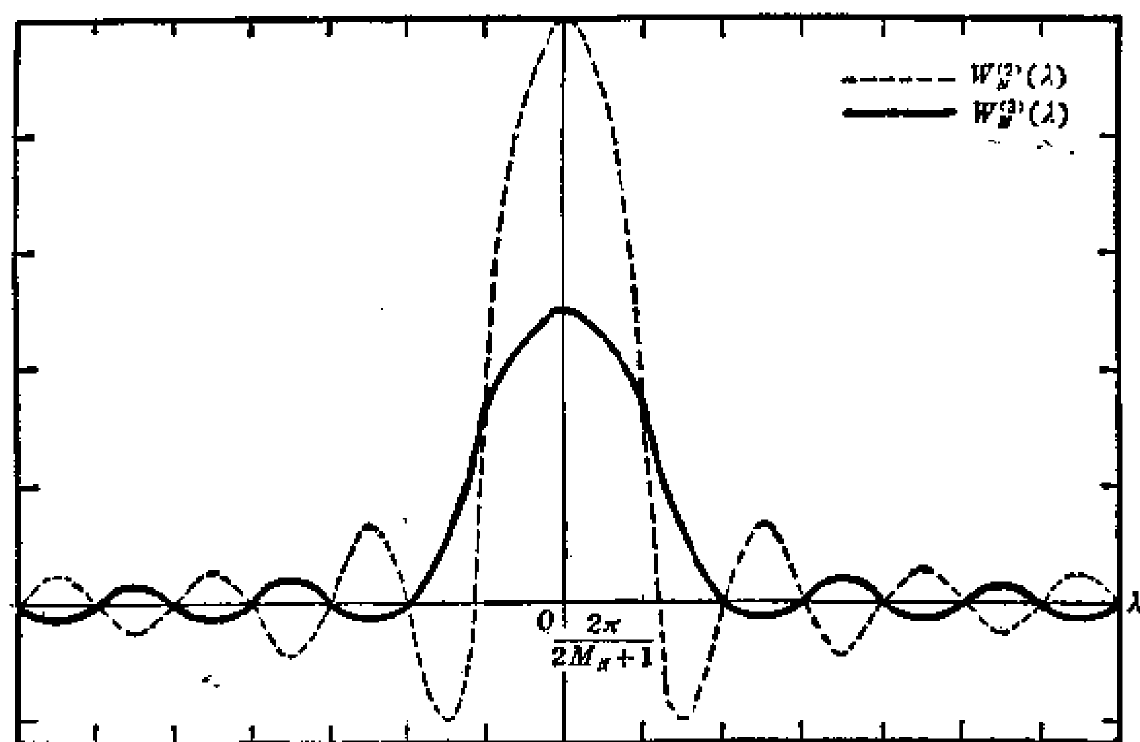


图 1

［例 3］ 若取

$$h(x)=\begin{cases}(1-\alpha)+\alpha\cos x, & |x|\leqslant 1\\ 0, & |x|>1\end{cases}$$

$$
w_{Nk}=\begin{cases}(1-\alpha)+\alpha\cos(B_N k), & |k|\leq\frac{1}{B_N}\\ 0, & |k|>\frac{1}{B_N}\end{cases}
$$

则取 $M_N=\left[\frac{1}{B_N}\right]$时,

$$
W_N^{(3)}(\lambda)=\frac{\alpha}{2}W_N^{(2)}\left(\lambda-\frac{\pi}{M_N}\right)+(1-\alpha)W_N^{(2)}(\lambda)
$$

$$
+\frac{\alpha}{2}W_N^{(2)}\left(\lambda+\frac{\pi}{M_N}\right) \tag{55}
$$

其中 $W_N^{(2)}(\lambda)$ 由(53)规定,而

$$
\hat{f}_N(\lambda)=\frac{1}{2\pi}\sum_{k=-M_N}^{M_N}\left[(1-\alpha)+\alpha\cos\left(\frac{\pi}{M_N}\right)\right]\hat{B}(k)e^{-ik\lambda}
$$

$$
=\int_{-\pi}^{\pi}W_N^{(3)}(\theta)I_N(\lambda-\theta)d\theta \tag{56}
$$

从(55)可见,$W_N^{(3)}(\lambda)$相当于把 $W_N^{(2)}(\lambda)$ 与其本身经过左右移动后的波形进行线性组合,由于移动的位置大约是$W_N^{(2)}(\lambda)$摆动的半个周期,所以移动又相加有消除$W_N^{(2)}(\lambda)$的边瓣的振动的目的. 当然也使$W_N^{(3)}(\lambda)$的主瓣不象$W_N^{(2)}(\lambda)$那样陡. 从计算上看(56)表明它也不需要计算全部 $\hat{B}(n)$, 所需的计算量和例 2 是相同的,所以在实用中这类估计量是经常被采用的,特别当

$\alpha=0.46$ 称为海明(Hamming)窗,

$\alpha=0.5$ 称为汉(Hann)窗.

这是最常用的两种情况. 在此因为$W_N^{(3)}(\lambda)$ 也可能取负值, 所以$\hat{f}_N(\lambda)$也可能取负值,这一缺陷,对于较为平滑的谱密度来说,不是严重的问题, 它的正负边瓣的影响在(56)的积分式中可以相互抵消. 但对于变化较大的谱密度或者对交互谱密度其本身也可能取负值,正负边瓣的影响在(56)的积分式有可能积累起来,这时就必须采用谱窗函数为正的估计量.

从 $h(x)$的取法还可看到，对这类窗当 $q=2$ 时，$h_q=\frac{\alpha}{2}$，因此当 $q\leqslant 2$，且 $\sum_{n=-\infty}^{\infty}|n|^q|B(n)|<\infty$ 时，定理 4 对这一类估计量也是适用的.

[例 4] 若取

$$h(x)=\begin{cases}1-6|x|^2(1-|x|), & |x|\leqslant\frac{1}{2}\\ 2(1-|x|)^3, & \frac{1}{2}\leqslant|x|\leqslant 1\\ 0, & |x|>1\end{cases}$$

$$w_{Nk}=h(B_Nk)$$

此时对应的$W_N^{(4)}(\lambda)$无简单的表示式，但可以说明它总是正的，而取 $M_N=\left[\frac{1}{B_N}\right]$时相应的估计量为

$$\hat{f}_N(\lambda)=\frac{1}{2\pi}\sum_{k=-M_N}^{M_N}w_{Nk}\hat{B}(k)e^{-ik\lambda}$$

$$=\int_{-\pi}^{\pi}W_N^{(4)}(\theta)I_N(\lambda-\theta)d\theta \tag{57}$$

对于这个估计量，它的谱窗函数$W_N^{(4)}(\lambda)$是恒正的，而且边瓣也比较小. 所以可以避免边瓣影响的积累，从计算上考虑，虽然 $\hat{B}(k)$的计算量是与例 3 中相同的，但 w_{Nk} 的计算要比前面稍复杂一点. 此外，从 $h(x)$的表示式可以看出，对 $q=2, h_q=6$，所以定理 4对它也可适用.

上面我们列举了一些谱密度估计量的例子，对于它们都可用定理 2 来考察其渐近均值和方差. 当然从减少偏差来说(49)表明B_N 趋于零快一点比较好，但是它的倒数$M_N=\left[\frac{1}{B_N}\right]$一般又是与需要计算 $\hat{B}(k)$的点数相联系的. 另一方面，从减小方差来说(50)

表明B_N趋于零慢一点好，假若要综合这两方面的要求，一种可行的方法是让

$$E|\hat{f}_N(\lambda)-f(\lambda)|^2=\min \tag{58}$$

由于

$$\begin{aligned}
&E|\hat{f}_N(\lambda)-f(\lambda)|^2\\
=&|E\{\hat{f}_N(\lambda)\}-f(\lambda)|^2+D\{\hat{f}_N(\lambda)\}\\
&\sim B_N^{2q}\left(\frac{h_q}{2\pi}\sum_{-\infty}^{\infty}|n|^q B(n)e^{-in\lambda}\right)^2\\
&+(NB_N)^{-1}f^2(\lambda)\int_{-\infty}^{\infty}h^2(x)dx
\end{aligned}$$

所以，从理论上说，B_N随N变化时，应满足下列关系式，

$$0<\varliminf_{N\to\infty}NB_N^{2q+1}\leqslant\varlimsup_{N\to\infty}NB_N^{2q+1}<\infty$$

才可能使(58)成立，当然在实际应用上(49)(50)右端涉及到的一些常数本身又是与谱密度$f(\lambda)$和协方差函数$B(n)$有关的．所以B_N的选择问题多半还是经验加试验来定的．

为了估计系统的频率特征，一般还需要估计输入、输出的交互谱对于交互谱密度，也可用类似于(44)的公式计算，只需在(44)中以互协方差函数估计量代替$\hat{B}(\nu)$即可，当然窗因子的选择应考虑到交互谱函数的特点，对于交互谱密度估计，不难说明定理4的结论也同样是成立的．

上面我们只是对时间序列$X(n)$来讨论它的谱密度估计问题，在实用上，虽然处理的是时间序列，但它却是某个连续参数过程的采样值，而处理的结果，又要求将它来说明连续参数过程的谱密度，为了做到这一点，先要把两者谱密度的关系推导出来，若连续参数过程为$\{X_c(t),t\in(0,\infty)\}$，它的采样值为

$$X_d(n)=X_c(n\Delta)$$

其中Δ为采样间隔，又$f_c(\lambda)$和$f_d(\lambda)$分别为$X_c(t)$和$X_d(n)$的谱

密度，假若采样间隔 Δ 是满足定理 4.3.2 的要求，那么

$$f_c(\lambda)=0, \qquad |\lambda|>\frac{\pi}{\Delta}$$

这时

$$\begin{aligned}\int_{-\pi}^{\pi} e^{ik\lambda} f_d(\lambda)d\lambda &= E\{X_d(n+k)\overline{X_d(n)}\} \\ &= E\{X_c((n+k)\Delta)\overline{X_c(n\Delta)}\} \\ &= \int_{-\frac{\pi}{\Delta}}^{\frac{\pi}{\Delta}} e^{ik\Delta\lambda} f_c(\lambda)d\lambda \\ &= \int_{-\pi}^{\pi} e^{ik\lambda}\Delta^{-1} f_c\left(\frac{\lambda}{\Delta}\right)d\lambda\end{aligned}$$

所以

$$f_d(\lambda)=\frac{1}{\Delta}f_c\left(\frac{\lambda}{\Delta}\right), \qquad |\lambda|\leqslant\pi$$

若 $\hat{f}_{dN}(\lambda)$ 为 $f_d(\lambda)$ 的估计量，那么

$$\hat{f}_{cN}(\lambda)=\Delta\hat{f}_{dN}(\lambda\Delta), \quad |\lambda|\leqslant\frac{\pi}{\Delta}$$

就可取为连续参数过程 $\{X(t)\}$ 的谱密度 $f_c(\lambda)$ 的估计量.

至此，我们只是介绍谱密度估计的基本内容，但是在实用中，还必须考虑它的计算方面的问题，因为习惯的做法是通过协方差函数估计量 $\hat{B}(n)$ 来计算谱密度的估计量 $\hat{f}_N(\lambda)$，而当我们获得了观测 $\{X(1),\cdots,X(N)\}$，每计算一个 $\hat{B}(k)$，其乘加计算次数是与 N 成比例的，为了进一步计算 $\hat{f}_N(\lambda)$，就要有大约 M_N 个 $\hat{B}(k)$，而对某一个 λ 计算 $\hat{f}_N(\lambda)$ 时，又要大约 M_N 次运算，而当我们对 k 个(一般与 M_N 有相同量级) λ 计算 $\hat{f}_N(\lambda)$ 时，总运算次数大约是

$$C_1NM_N+C_2M_N^2$$

当 N, M_N 很大时，这是一个不可忽视的运算量，尤其因为谱分析在时间序列分析中是最基本的手段之一，它的运算方面就更要引

起重视了．在六十年代出现的快速富利埃变换算法大大地减轻了谱分析中计算的负担，使它有可能作为时间序列分析的基本工具被广泛地采用．快速富利埃变换又称 FFT，是指在计算下列离散富利埃变换时

$$A(k)=\sum_{j=0}^{M-1}a_jW^{jk},\quad W=e^{-i\frac{2\pi}{M}},\quad 0\leqslant K\leqslant M-1$$

FFT 算法可以把原来要进行 M^2 次运算获得的结果用 $M\log_2 M$ 次运算来求得，特别当 $M=2^N$ 时，这一算法在数字电子计算机上实现又极为方便，所以它的出现大大推动了谱分析的应用与发展．它不仅减少了 $\hat{B}(k)$ 求 $\hat{f}_N(\lambda)$ 时的计算量，而且利用"函数卷积的富利埃变换是函数富利埃变换的乘积"这一性质，在协方差函数的计算中也采用 FFT 算法，把 C_1NM_N 次运算减少到大约 $C_2N\log_2 M_N$ 次运算，不仅如此，还出现了形如

$$\hat{f}_N(\lambda)=\frac{1}{2\pi}\left|\sum_{k=1}^{N}X(k)\,w_{Nk}e^{-ik\lambda}\right|^2$$

的估计量和直接由周期图加窗的种种算法，这些都是利用了 FFT 算法在计算上的有效性，也大大丰富了谱分析的内容，由于这些都超出本书的范围，所以我们不去介绍它了．

§4. 自回归模型拟合

上一节我们介绍了时间序列的二阶统计特征——协方差函数和谱密度的估计问题，对于时间序列我们求出它的二阶统计特征是出于不同的实际需要．在分析线性系统时，由系统的输入、输出的二阶统计特性可以求得系统的频率特性，但在另一些问题，求得时间序列的二阶统计特征是为了模拟产生这一时间序列或者对时间序列进行预报．例如把路面的颠簸不平变化看为一个时间序列

时，获得这个时间序列的统计特征就是为了在研究汽车的振动分析中模拟发生这种路面的不平．在水文、气象中测量到的一些时间序列里，掌握它的统计特征是为了便于进行预报．从模拟和预报的要求来看，假定时间序列本身满足某个物理或数学的模型是方便的，因为有了物理模型，便于由它的机理推出它的数学模型，而有了数学的模型，就便于进行模拟和预报．

在考虑描述时间序列的各类模型时，上一章提到的滑动和自回归混合模型就成为最基本的一类模型，因为满足 ARMA 过程可看为白噪声通过差分方程描述的线性时不变系统而发生的，所以它能近似地符合大多数线系统中噪声的模型．从数学上看，它在平稳序列的情况下是具有有理谱密度的，是可以近似任何具有连续谱密度的平稳过程的．从预报上看，§1 中已表明对 ARMA 过程的预报问题解决是较为方便的，尤其是对自回归过程．而下一章将表明对 ARMA 过程的模拟也是方便的，所以我们要来讨论用 ARMA 模型来拟合时间序列的问题．在此，当然包括从时间序列的物理模型来看是已满足某个 ARMA 模型的情形，也包括用 ARMA 模型去近似地描述某个时间序列的情况，限于数学处理上的方便，本书主要只讨论 ARMA 模型的拟合问题．

在这一节中，为了简单起见，我们假定被观测的时间序列 $\{Z(n)\}$ 是均值为零的平稳序列．即在(3.1)中，

$$m(n)=0, \qquad Z(n)=X(n)$$

否则，只要 $\{Z(n)\}$ 满足(3.1)——(3.3)，我们也可以用 §3 的办法来处理．

一、自回归模型的参数估计问题

若 $\{X(n)\}$ 为满足下列自回归模型的平稳序列

$$\sum_{j=0}^{p} a_j X(n-j)=\xi(n), \quad a_0=1 \tag{1}$$

其中$\{\xi(n)\}$为白噪声序列，则当a_j满足下列条件时，

$$\sum_{j=0}^{p} a_j \lambda^{p-j}=0 \text{ 的根的模全小于 } 1 \tag{2}$$

由定理 4.4.5 的系可知

$$\xi(n) \perp H_{n-1}(X) \tag{3}$$

又由定理 4.4.2 的系可知，这时$X(n)$的协方差函数$B(n)$满足下列尤尔-瓦尔克方程

$$\sum_{j=0}^{p} a_j B(n-j)=0,\quad n\geqslant 1,\quad a_0=1 \tag{4}$$

而且利用(3)和$a_0=1$，可得

$$\begin{aligned}\sigma^2 &\triangleq E|\xi(n)|^2\\ &=E\left|\sum_{j=0}^{p} a_j X(n-j)\right|^2\\ &=\sum_{k=0}^{p}\bar{a}_k\left(\sum_{j=0}^{p} a_j B(k-j)\right)\\ &=\sum_{j=0}^{p} a_j B(-j)\end{aligned} \tag{5}$$

这便是我们拟合自回归模型的出发点，对时间序列$\{X(n)\}$先假定它是满足(1)，若由$\{X(n)\}$观测值求得$B(n)$的如下估计值

$$\hat{B}(k)=\frac{1}{N}\sum_{n=1}^{N-|k|} X(n)X(n+|k|) \tag{6}$$

则由$\{\hat{B}(k),0\leqslant k\leqslant p\}$便可按(4)求出$a_j$的估计量$\hat{a}_j$，并由(5)求得$\sigma^2$的估计量$\hat{\sigma}^2$. 若将求解$\hat{a}_j$的方程写出，它便是

$$\sum_{j=1}^{p} a_j \hat{B}(i-j)=-\hat{B}(i),\quad 1\leqslant i\leqslant p \tag{7}$$

上述方程组的系数阵为

$$\Gamma \triangleq \begin{bmatrix} \hat{B}(0) & \hat{B}(-1) & \cdots & \hat{B}(-p+1) \\ \hat{B}(1) & \hat{B}(0) & \cdots & \hat{B}(-p+2) \\ \cdots & \cdots & \cdots & \cdots \\ \hat{B}(p-1) & \hat{B}(p-2) & \cdots & \hat{B}(0) \end{bmatrix} \tag{8}$$

只要 $\hat{B}(0)\neq 0$，即当诸 $X(n)$ 不全为零时，上述 Γ 就是正定的(见习题)，这样满足(7)的 $\hat{a}_1, \hat{a}_2, \cdots, \hat{a}_p$ 是唯一的。
此外，若假定当 $n\leqslant 0, X(n)=0$，则由

$$\sum_{n=1}^{N}\left|\sum_{j=0}^{p}a_j X(n-j)\right|^2 = \min$$

求得的关于 a_j 的方程式也类似于(7)，所以由(7)求得的估计也近似于"最小二乘方"估计.

对于(8)中的矩阵 Γ，它与主对角线平行的斜线上的元素都是相同的，它又是非负定的，这类矩阵又称托布里兹矩阵．而以正定托布里兹阵为系数阵的线性方程组，其求解有专门的递推算法，这种算法的运算次数是与其阶数 p 平方成比例，较一般消去法为快，所以对(7)的求解在实用上是十分方便的.

为了证明由(7)求得的 $\hat{a}_j$ 的渐近性质，我们要用到下列引理.

引理 1 若 p 维随机变量 ξ_n 的分布收敛于 $N(0, \Sigma)$ 分布，又

$$\eta_n = U_n + V_n\xi_n$$

其中 p 维随机变量 U_n 的每个分量都依概率收敛于零，而 $p\times p$ 的随机变量构成的矩阵 V_n 的每个分量依概率收敛于常数矩阵 A 的每个分量，则 p 维随机变量 η_n 的分布收敛于 $N(0, A\Sigma A')$ 分布.

[证明] 为叙述简单起见，我们只对 $p=1$ 来进行证明，若 ξ_n 的极限分布为 $N(0, \sigma^2)$，V_n 依概率收敛于 $A>0$，则对任一 $\varepsilon>0$，

$$\begin{aligned} P(\eta_n<x) &\leqslant P(|U_n|\geqslant\varepsilon)+P(V_n\xi_n<x+\varepsilon) \\ &\leqslant P(|U_n|\geqslant\varepsilon)+P(|V_n-A|\geqslant\varepsilon)+P\left(\xi_n<\frac{x+\varepsilon}{A-\varepsilon}\right) \end{aligned}$$

所以

$$\overline{\lim_{n\to\infty}} P(\eta_n<x)\leqslant\lim_{n\to\infty} P\left(\xi_n<\frac{x+\varepsilon}{A-\varepsilon}\right)$$

$$=\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\frac{x+\varepsilon}{A-\varepsilon}} e^{-\frac{u^2}{2\sigma^2}}du$$

由 ε 的任意性，令 $\varepsilon\to0$ 即得

$$\overline{\lim_{n\to\infty}} P(\eta_n<x)\leqslant\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\frac{x}{A}} e^{-\frac{u^2}{2\sigma^2}}du$$

$$=\frac{1}{\sqrt{2\pi}A\sigma}\int_{-\infty}^{x} e^{-\frac{u^2}{2A^2\sigma^2}}du \tag{9}$$

同理，对任一 $\varepsilon>0$ 有

$$P(\eta_n<x)\geqslant -P(|U_n|\geqslant\varepsilon)+P(V_n\xi_n<x-\varepsilon)$$

$$\geqslant -P(|U_n|\geqslant\varepsilon)-P(|V_n-A|\geqslant\varepsilon)$$

$$+P\left(\xi_n<\frac{x-\varepsilon}{A+\varepsilon}\right)$$

所以

$$\underline{\lim_{n\to\infty}} P(\eta_n<x)\geqslant\lim_{n\to\infty} P\left(\xi_n<\frac{x-\varepsilon}{A+\varepsilon}\right)$$

$$=\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\frac{x-\varepsilon}{A+\varepsilon}} e^{-\frac{u^2}{2\sigma^2}}du$$

由 ε 的任意性，令 $\varepsilon\to0$ 可得

$$\underline{\lim_{n\to\infty}} P(\eta_n<x)\geqslant\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\frac{x}{A}} e^{-\frac{u^2}{2\sigma^2}}du$$

$$=\frac{1}{\sqrt{2\pi}A\sigma}\int_{-\infty}^{x} e^{-\frac{u^2}{2A^2\sigma^2}}du \tag{10}$$

比较(9)，(10)即得

$$\lim_{n\to\infty} P(\eta_n<x)=\frac{1}{\sqrt{2\pi}A\sigma}\int_{-\infty}^{x} e^{-\frac{u^2}{2A^2\sigma^2}}du$$

所以引理对 $p=1$ 及 $A>0$ 时是成立的，对 $A<0$ 或多维情况也可类似地加以证明.

定理 1 若平稳正态过程 $\{X(n)\}$ 满足自回归模型(1)，(2)，$\hat{B}(k)$ 由(6)给出，而 a_j 的估计量 $\hat{a}_j$，$1\leqslant j\leqslant p$ 由(7)给出，则 $\hat{a}_j$，$1\leqslant j\leqslant n$ 具有下列性质:

(i) $\hat{a}_j$ 是 a_j 的一致性估计，即对任一 $\varepsilon>0$

$$\lim_{N\to\infty}P(|\hat{a}_j-a_j|>\varepsilon)=0,\quad 1\leqslant j\leqslant p$$

(ii) 若 $\sigma^2=E\xi^2(n)$，又

$$\hat{\sigma}_1^2=\frac{1}{N}\sum_{n=p+1}^{N}(X(n)+\sum_{j=1}^{p}\hat{a}_jX(n-j))^2 \tag{11}$$

$$\hat{\sigma}_2^2=\sum_{j=0}^{p}\hat{a}_j\hat{B}(-j)$$

则 $\hat{\sigma}_1^2,\hat{\sigma}_2^2$ 都是 σ^2 的一致性估计.

(iii) $\hat{a}_1,\cdots,\hat{a}_p$ 是渐近正态的，即

$$\sqrt{N}(\hat{a}_1-a_1),\ \sqrt{N}(\hat{a}_2-a_2),\cdots,\sqrt{N}(\hat{a}_p-a_p)$$

的分布收敛为 $N(0,\Sigma)$ 分布，而

$$\Sigma=(\sigma_{kl})=\sigma^2(B(k-l))^{-1} \tag{12}$$

[证明] 在证明 $\hat{a}_j$ 和 $\hat{\sigma}^2$ 的一致性时主要用到第二册的定理 4.5.4，即若 $f(x)$ 为连续函数，则当 $n\to\infty$ 时，若

$$\xi_n\xrightarrow{P}\xi$$

便有
$$f(\xi_n)\xrightarrow{P}f(\xi)$$

因为 $\{X(n)\}$ 有有理谱密度，$(B(k-l))$ 非奇异，所以由(4)

$$\begin{bmatrix}a_1\\ \vdots\\ a_p\end{bmatrix}=-(B(k-l))^{-1}\begin{bmatrix}B(1)\\ \vdots\\ B(p)\end{bmatrix}$$

而由(7)

$$\begin{bmatrix}\hat{a}_1\\ \vdots \\ \hat{a}_p\end{bmatrix} = -(\hat{B}(k-l))^{-1}\begin{bmatrix}\hat{B}(1)\\ \vdots \\ \hat{B}(p)\end{bmatrix}$$

又由定理 3.2, $\hat{B}(k), 0\leqslant k\leqslant p$ 是 $B(k)$的一致性估计，所以由第二册的定理 4.5.4 $\hat{a}_1, \cdots, \hat{a}_p$ 就是 $a_1, \cdots, a_p$ 的一致性估计. 同样, 利用(5)不难推出 $\hat{\sigma}_1^2, \hat{\sigma}_2^2$ 都是由 σ^2 的一致性估计.

为了证明 $\hat{a}_j$ 的渐近正态性, 首先令

$$U_N^{(i)} \triangleq \sum_{j=0}^{p}\hat{a}_j\left\{\frac{1}{N}\sum_{n=p}^{N}X(n-j)X(n-i)\right\}, a_0=1, 1\leqslant i\leqslant p \tag{13}$$

又由(7)

$$\sum_{j=0}^{p}\hat{a}_j\hat{B}(i-j)=0, \quad \hat{a}_0=1, \quad 1\leqslant i\leqslant p$$

把上两式相减不难推出

$$p-\lim_{N\to\infty}\sqrt{N}U_N^{(i)}=0, \quad 1\leqslant i\leqslant p$$

又由(1)

$$\begin{aligned}\xi_N^{(i)} &\triangleq \sum_{j=0}^{p}a_j\left\{\frac{1}{N}\sum_{n=p}^{N}X(n-j)X(n-i)\right\}\\ &=\frac{1}{N}\sum_{n=p}^{N}\left(\sum_{j=0}^{p}a_jX(n-j)\right)X(n-i)\\ &=\frac{1}{N}\sum_{n=p}^{N}\xi(n)X(n-i)\end{aligned} \tag{14}$$

因此 $\xi_N^{(i)}$ 可看为$\{\xi(n)\}$, $\{X(n)\}$ 协方差函数 $B_{\xi X}(i)$的估计量, 因为(3)成立, 所以

$$B_{\xi X}(n)=0, \quad n\geqslant 1$$

而

$$B_{\xi\xi}(n)=\sigma^2\delta_{n0}$$

因此由定理 3.4,

$$\sqrt{N}\xi_N^{(1)}, \sqrt{N}\xi_N^{(2)}, \cdots, \sqrt{N}\xi_N^{(p)}$$

是渐近正态 $N(0,\Sigma)$ 分布的，而

$$\Sigma=(\sigma_{kl})=\sigma^2(B_{XX}(k-l))=\sigma^2(B(k-l)).$$

把(13)，(14)相减可得

$$\sum_{j=1}^{p}\sqrt{N}(\hat{a}_j-a_j)\left\{\frac{1}{N}\sum_{n=p}^{N}X(n-j)X(n-i)\right\}$$
$$=\sqrt{N}U_N^{(i)}-\sqrt{N}\xi_N^{(i)}$$

若记

$$V_{ij}(N)=\frac{1}{N}\sum_{N=p}^{N}X(n-j)X(n-i)$$
$$V_N=(V_{ij}(N))$$

则由定理 3.2，$V_{ij}(N)$ 依概率收敛于 $B(i-j)$，V_N 依概率收敛于 $(B(i-j))$，而

$$\begin{bmatrix}\sqrt{N}(\hat{a}_1-a_1)\\ \vdots\\ \sqrt{N}(\hat{a}_p-a_p)\end{bmatrix}=V_N^{-1}\begin{bmatrix}\sqrt{N}U_N^{(1)}\\ \vdots\\ \sqrt{N}U_N^{(p)}\end{bmatrix}-V_N^{-1}\begin{bmatrix}\sqrt{N}\xi_N^{(1)}\\ \vdots\\ \sqrt{N}\xi_N^{(1)}\end{bmatrix}$$

所以由引理 1 可得

$$\sqrt{N}(\hat{a}_1-a_1), \cdots, \sqrt{N}(\hat{a}_p-a_p)$$

的联合分布是收敛于正态分布的，且其极限分布的协方差阵为(12). 定理证毕.

上面我们只是介绍了自回归模型的参数 a_j, $1\leqslant j\leqslant p$ 和 σ^2 的一种估计量及其在 $X(n)$ 为正态过程时的渐近性质. 事实上，当 $\{X(n)\}$ 为正态过程时，由(7)，(11) 求得的 $\hat{a}_j$ 和 $\hat{\sigma}^2$ 是与最大似然估计渐近相同的. 为了说明这一点只需写出 $X(1), X(2), \cdots, X(N)$ 的分布密度即可看出，因为

$$\xi(n)=\sum_{j=0}^{p}a_jX(n-j),\ p+1\leqslant n\leqslant N,\ a_0=1$$

所以，若记 $X(1),\cdots,X(p)$ 的分布密度为 $f_p(x_1,\cdots,x_p)$，则 $X(1),\cdots,X(N)$ 的分布密度为

$$
\begin{aligned}
& f_{X(1)\cdots X(N)}(x_1,\cdots,x_N) \\
=& f_{X(1)\cdots X(p)\xi(p+1)\cdots\xi(N)}(x_1,\cdots,x_p,y_1,\cdots,y_N) \\
=& f_{X(1)\cdots X(p)}(x_1,\cdots,x_p)\prod_{n=p+1}^{N} f_{\xi(n)}(y_n) \\
=& f_p(x_1,\cdots,x_p)\prod_{n=p+1}^{N}\frac{1}{\sqrt{2\pi}\,\sigma}e^{-\frac{y_n^2}{2\sigma^2}} \\
=& f_p(x_1,\cdots,x_p)(\sqrt{2\pi}\,\sigma)^{-(N-p)}\exp\left(-\frac{1}{2\sigma^2}\sum_{n=p+1}^{N}y_n^2\right) \\
=& f_p(x_1,\cdots,x_p)(\sqrt{2\pi}\,\sigma)^{-(N-p)}\exp\left(-\frac{1}{2\sigma^2}\sum_{n=p+1}^{N}\left(\sum_{j=0}^{p}a_jx_{n-j}\right)^2\right)
\end{aligned}
$$

由于 $f_p(x_1,\cdots,x_p)$ 是与 N 无关的正态分布密度，所以由

$$
(\sqrt{2\pi}\,\sigma)^{-(N-p)}\exp\left(-\frac{1}{2\sigma^2}\sum_{n=p+1}^{N}\left(\sum_{j=0}^{p}a_jx_{n-j}\right)^2\right)=\max \tag{15}
$$

求得的 $\hat{a}$，$\hat{\sigma}^2$ 与由 $f_{X(1)\cdots X(N)}(x_1,\cdots,x_p)=\max$ 求得的估计量，即最大似然估计，当 N 趋于无穷时是渐近等效的，而由 (15) 求得的 $\hat{a}_j$，$\hat{\sigma}^2$ 又与 (7)，(12) 求得的 $\hat{a}_j$，$\hat{\sigma}^2$ 是渐近相同的.

二、自回归模型的拟合

上面我们介绍了自回归模型中参数的估计问题，在那里是假定了 $\{X(n)\}$ 本身是满足自回归模型的，为了预报和控制的需要常常要对时间序列 $\{X(n)\}$ 用特定的模型去拟合，这时对 $\{X(n)\}$ 就并不假定它是否满足某种模型，而是用拟合的模型作为近似描述 $\{X(n)\}$ 统计规律性的一种工具，有了这种近似的统计规律性在预报和控制上都是方便的. 下面我们就来考虑用自回归模型去拟合时间序列的问题.

若被观测的时间序列为 $\{X(n)\}$，自回归模型拟合就是要确定

一个"合适的"的自回归模型，使满足这一模型平稳过程$\{X_1(n)\}$与$\{X(n)\}$有相近的统计规律性．从上面的讨论已经知道，要确定一个p阶自回归模型实际上就是定出它的参数$a_1, \cdots, a_p, \sigma^2$．这时，不论$\{X(n)\}$本身是否满足自回归模型，上面介绍的估计$a_1, \cdots, a_p$和$\sigma^2$的方法仍然是可以用的，即由$X(n)$仍可以算出其样本的协方差函数

$$\hat{B}(k)=\frac{1}{N}\sum_{n=1}^{N-|k|}X(k+|n|)X(k), \quad 0\leqslant k\leqslant p \tag{16}$$

并由$\hat{B}(k)$构成下列方程组

$$\sum_{j=1}^{p}a_j\hat{B}(k-j)=-\hat{B}(k), \quad 1\leqslant k\leqslant p \tag{17}$$

同前面一样的理由可由(17)解出$a_1, \cdots, a_p$，再按

$$\sigma^2=\hat{B}(0)+\sum_{j=1}^{p}a_j\hat{B}(-j) \tag{18}$$

求得σ^2，由(16)——(18)解出的$a_1, \cdots, a_p, \sigma^2$便可确定一个$p$阶自回归模型

$$\left.\begin{array}{l}\sum_{j=0}^{p}a_jX_1(n-j)=\xi_1(n), a_0=1\\ E\{\xi_1(n)\xi_1(m)\}=\delta_{nm}\sigma^2\end{array}\right\} \tag{19}$$

(19)就可作为近似描述$\{X(n)\}$统计规律性的模型．为了说明(19)中$X_1(n)$的一些统计性质，我们需要下列结果．

引理2 若(17)的系数矩阵$(\hat{B}(k-j))$为正定的，则它存在唯一解$a_1, \cdots, a_p$，且

$$\lambda^p+\sum_{j=1}^{p}a_j\lambda^{p-j}=0 \text{ 的根全在圆 } |\lambda|<1 \text{ 之内} \tag{20}$$

引理3 若$a_1, \cdots, a_p$满足(20)，不论$\{X(n)\}$是否满足自回归模型，满足

$$B(k)+\sum_{j=1}^{p}a_jB(k-j)=0,\quad 1\leqslant k\leqslant p \tag{21}$$

的自相关序列 $R(k)=\dfrac{B(k)}{B(0)}$, $0\leqslant k\leqslant p$ 是唯一的.

由于这两个引理的证明要用到较多的分析知识，我们在此把它的证明省略了①.

引理 2 表明，不论$\{X(n)\}$是否满足自回归模型，我们用(16)——(18)的做法得到的自回归模型总满足(20)，这正是§1考虑预测问题时所需要的一种模型，所以用这一办法拟合到的自回归模型可直接用于预报和控制.

由引理 3 立即可推出

定理 2 若$\{B_1(n)\}$为(19)中$\{X_1(n)\}$的协方差函数,则

$$B_1(n)=\hat{B}(n),\quad 0\leqslant n\leqslant p \tag{22}$$

[证明] 由引理 2, a_j 满足(20). 另一方面，由定理 4.5.3 的系可知,(19)中的 $X_1(n)$的谱密度 $f_1(\lambda)$ 可表为

$$f_1(\lambda)=\frac{\sigma^2}{2\pi\left|\sum\limits_{j=0}^{p}a_je^{-ij\lambda}\right|^2},\quad a_0=1 \tag{23}$$

所以对$\{X_1(n)\}$的协方差函数 $B_1(n)$,当 $k\geqslant1$ 时有

$$\begin{aligned}&\sum_{j=0}^{p}a_jB_1(k-j)\\&=\int_{-\pi}^{\pi}e^{ik\lambda}\left(\sum_{j=0}^{p}a_je^{-ij\lambda}\right)f_1(\lambda)d\lambda\\&=\frac{\sigma^2}{2\pi}\int_{-\pi}^{\pi}\frac{e^{ik\lambda}}{\sum\limits_{j=0}^{p}\overline{a_j}e^{ij\lambda}}d\lambda\end{aligned}$$

---

① 参见 J. Makhoul, Proc. IEEE, Vol. 63, 561——580 (1975) 及其所引的文献.

$$=\frac{\sigma^2}{2\pi i}\oint\frac{z^{k-1}dz}{\sum_{j=0}^{p}\overline{a_j}z^j}=0 \tag{24}$$

上述等式的最后一步是因为$\sum_{j=0}^{p}\overline{a_j}z^j=0$的根全在圆外，所以当$k\geqslant 1$时被积函数在圆内解析，在圆周上积分为零。

而当$k=0$时，因为$a_0=1$，

$$\sum_{j=0}^{p}a_jB_1(-j)$$

$$=\frac{\sigma^2}{2\pi i}\oint\frac{dz}{z\left(\sum_{j=0}^{p}\overline{a_j}z^j\right)}=\sigma^2 \tag{25}$$

由(24)，$B_1(k)$也满足(21)，按引理 3

$$\frac{\hat{B}(n)}{\hat{B}(0)}=\frac{B_1(n)}{B_1(0)},\quad 0\leqslant n\leqslant p$$

即

$$B_1(n)=\frac{B_1(0)}{\hat{B}(0)}\hat{B}(n),\quad 0\leqslant n\leqslant p$$

又由上式及(18)，(25)可知

$$B_1(n)=\hat{B}(n),\quad 0\leqslant n\leqslant p$$

定理证毕。

上述定理表明用(16)——(18)的方法去拟合p阶自回归时，是保证了用以拟合的$X_1(n)$的协方差函数和$X(n)$的样本的协方差函数其前$p+1$个值是相符合的，这与统计中的矩方法是十分相象的，由于(16)——(18)的做法方便及引理 2 和定理 2 表明的理由，这是一种最常用的方法。

在实际运用自回归模型去拟合时间序列资料时，还常发生这样的问题，即自回归模型的阶p该取多少为好，当然我们有事先的

经验提供参考，这是最好的，假若无这方面的精确知识，就应该从实际资料中估计一个 p，对于这一问题，还无定型的方法．一种做法是从假设检验的思想出发，先对 $p+h$，估计出 $\hat{a}_k$，$1\leqslant k\leqslant p+h$，然后来检验

$$H_0:\ a_{p+1}=a_{p+2}=\cdots=a_{p+h}=0$$

它可以用 $\hat{a}_k$ 的渐近分布来进行严格的检验，也可以从 a_k 的估计值 $\hat{a}_k$ 直观地定出一个 p，但是这种做法受到不少非议，一方面，因为产生时间序列的模型往往是十分复杂的，它一般并不满足自回归模型，而自回归模型只是用来拟合实际模型的近似式，所以缺乏进行假设检验的逻辑基础．另一方面，对实际资料运用模拟试验表明这个做法也是不很成功．所以有人提出另一种想法，即应该把模型拟合的问题当作 p 的估计问题来处理，但这时参数估计中常用的最小二乘法估计（或在正态时相当的最大似然估计）已是不适用了，因为用最小二乘法求 p，就是要使

$$\sum_{n=1}^{N}\left(\sum_{j=0}^{p}a_jX(n-j)\right)^2=\min,\quad a_0=1$$

来求 p 和 a_k，而在现在的情况下，随着 p 的增大，上述 $\min\limits_{a_j}\sum_{n=1}^{N}\left(\sum_{j=0}^{p}a_jX(n-j)\right)^2$ 就变小，所以是无法对 p 求得一个极值的，过大的 p，固然能使 $\sum_{n=1}^{N}\left(\sum_{j=0}^{p}\hat{a}_jX(n-j)\right)^2$ 很小，但把这一模型用于预报时，过多的 $\hat{a}_j$ 反而会带来预报误差．所以从模型拟合的要求来估计 p，必须提出新的准则，一个讨论较多的准则是用

$$I(p)=\log\sigma_p^2+\frac{2p}{N}=\min \tag{26}$$

去找 p，其中

$$\hat{\sigma}_p^2=\frac{1}{N}\sum_{n=1}^{N}\left(\sum_{j=0}^{p}\hat{a}_jX(n-j)\right)^2,\quad a_0=1 \tag{27}$$

$$X(l)=0,\quad l<0$$

从(26), (27)可见, $\hat{\sigma}_p^2$ 随 p 增大而减小, 而 $2p/N$ 随 p 增大而增大, 所以 $I(p)$ 有可能在某个 p 使 $I(p)$ 达到极小, 模拟试验的结果表明这种做法在实用中还是可取的①.

三、自回归模型拟合与谱密度估计

上面我们介绍了用自回归模型去拟合时间序列 $\{X(n)\}$ 的做法. 这一做法也可从频率域来加以考虑. 若被拟合的序列 $\{X(n)\}$ 有谱密度 $f(\lambda)$, 则在定理 2 的证明中已提到用以拟合的(19)中的 $\{X_1(n)\}$, 它的谱密度 $f_1(\lambda)$ 可表为

$$f_1(\lambda)=\frac{\sigma^2}{2\pi\left|\sum_{j=0}^{p}a_je^{-ij\lambda}\right|^2},\quad a_0=1 \tag{28}$$

因为 a_j, σ^2 都是根据 $X(n)$ 的样本按(16)——(18)求出的, 所以 $f_1(\lambda)$ 也可看为 $f(\lambda)$ 的一个估计量, 因此模型拟合问题在频率域上也是一个谱的估计问题, 不过在这里选择的估计量是形式完全确定而只依赖于有限个参数的那一类估计量, 所以是一种有限参数模型的谱密度估计问题.

在§3 曾提到谱密度的周期图估计量

$$I_N(\lambda)=\frac{1}{2\pi N}\left|\sum_{n=1}^{N}X(n)e^{-in\lambda}\right|^2$$

假若我们把它当作某个谱密度, 对应的协方差函数 $B_1(n)$ 应该是

$$\begin{aligned}B_1(k)&=\int_{-\pi}^{\pi}e^{ik\lambda}I_N(\lambda)d\lambda\\&=\frac{1}{N}\sum_{j=0}^{N-|k|}X(j+|k|)X(j)\\&=\hat{B}(k),\quad |k|\leqslant N\end{aligned} \tag{29}$$

① 关于这方面的讨论可参见 H. Akaike, IEEE Trans. AC-19, (1974)716——723; E. Parzen, IEEE Trans. AC-19, (1974)723——730.

$$B_1(k)=\int_{-\pi}^{\pi} e^{ik\lambda} I_N(\lambda)\,d\lambda=0, \quad |k|>N \tag{30}$$

在§3也曾提到周期图估计的不足之点，那里采用窗函数的办法来弥补周期图的不足，若采用的估计量 $f_N(\lambda)$ 为

$$f_N(\lambda)=\frac{1}{2\pi}\sum_{k=-N}^{N}\hat{B}(k)w_{Nk}e^{-ik\lambda}$$

则相应的协方差序列 $B_1(n)$ 为

$$\begin{aligned} B_1(n) &= \int_{-\pi}^{\pi} e^{in\lambda} f_N(\lambda)\,d\lambda \\ &= \begin{cases} \hat{B}(k)w_{Nk}, & |k|\leqslant N \\ 0, & |k|>N \end{cases} \end{aligned}$$

所以加窗的做法是放弃了(29)，保留了(30)来寻求较好的谱密度估计量．若采用(28)中的 $f_1(\lambda)$ 来估计 $f(\lambda)$，那么由定理2，

$$\begin{aligned} B_1(k) &= \int_{-\pi}^{\pi} e^{ik\lambda} f_1(\lambda)\,d\lambda \\ &= \hat{B}(k), \quad |k|\leqslant p \end{aligned}$$

而利用自回归模型协方差函数满足尤尔-瓦尔克方程(4.4.26)，不难推出 $|k|>p$ 时，$B_1(k)$ 不全为零，所以从谱估计来看，这里的做法是保留了(29)而放弃了(30)，(28)中的 $f_1(\lambda)$ 作为一个谱密度估计量，从理论上来讨论它的一些统计特性，这方面的结果还不多，但是已有不少模拟试验表明，这一估计方法在实用上具有短时，高分辨率的优点，所谓短时是指这一方法相对于§3中的方法，可以用较短的资料按(16)——(18)的做法，把谱密度的基本形态估计出来，所以它就适用于这样一类信号，它们从局部来看，可近似地当为平稳时间序列来处理，对这类信号谱密度估计的短时性就很重要．所谓高分辨率是指这种估计法当 $f(\lambda)$ 本身存在若干突峰时，$f_1(\lambda)$ 能较敏感的反映出来．因为在§3中介绍的加窗的做法，往往起到平滑的作用，使估计量 $f_N(\lambda)$ 在突峰附近不会很陡．而

(28)中的 $f_1(\lambda)$ 由于其函数形式的有利，只要分母上$\sum_{j=0}^{p}a_jZ^{p-j}$的零点靠近$e^{i\lambda_0}$，就可以使 $f_1(\lambda)$ 在λ_0附近很陡，而且做到这一点未必要求N一定是非常大. 正由于这样，这种谱密度的估计方法在语音分析和地球物理资料分析中获得了成功，受到了人们的注意. 下面我们介绍这种估计法的某些极值性质.

性质 1 与(28)中$f_1(\lambda)$所对应的协方差函数是满足

$$B_1(n)=\hat{B}(n), \quad |n|\leqslant p \tag{31}$$

而使

$$\det\Gamma_{p+1}=\det\begin{bmatrix}B_1(0) & \cdots B_1(-p-1)\\ \cdots\cdots & \cdots\cdots\\ B_1(p+1) & \cdots B_1(0)\end{bmatrix}$$

为最大者.

［证明］ 为使$\det\Gamma_{p+1}=\max$，必须有$\frac{\partial(\det\Gamma_{p+1})}{\partial B_1(p+1)}=0$，利用$B_1(n)=B_1(-n)$可得

$$\frac{\partial(\det\Gamma_{p+1})}{\partial B_1(p+1)}$$

$$=(-1)^{p+3}\begin{vmatrix}B_1(1) & B_1(0) & B_1(-1) & \cdots B_1(-p+1)\\ B_1(2) & B_1(1) & B_1(0) & \cdots B_1(-p+2)\\ \cdots\cdots & \cdots\cdots & \cdots\cdots & \cdots\cdots\\ B_1(p+1) & B_1(p) & B_1(p-1) & \cdots B_1(1)\end{vmatrix} \tag{32}$$

所以当$B_1(n)$是$f_1(\lambda)$所对应的协方差函数，则由定理4.4.3，$B_1(n)$满足尤尔-瓦尔克方程

$$B_1(k)=\sum_{j=1}^{p}a_jB_1(k-j), \quad 1\leqslant k\leqslant p$$

这就表明(32)右端行列式中第一列是其后p列的线性组合，所以

$$\frac{\partial(\det\Gamma_{p+1})}{\partial B_1(p+1)}=0$$

另一方面

$$\frac{\partial^2(\det\Gamma_{p+1})}{\partial B_1(p+1)^2}$$

$$=-\begin{vmatrix} B_1(0) & \cdots & B_1(-p+1) \\ \cdots & \cdots & \cdots \\ B_1(p-1) & \cdots & B_1(0) \end{vmatrix}<0$$

所以由 $f_1(\lambda)$ 所对应的 $B_1(p+1)$ 使 $\det\Gamma_{p+1}=\max$.

同样可以证明. 若在固定 $B_1(p+1)$ 取为 $f_1(\lambda)$ 所对应的协方差函数，那么 $B_1(p+2)$ 又使 $\det\Gamma_{p+2}=\max$. 并且可以一直递推下去. 在某些信息论问题中，对具有密度 $p(x)$ 的分布，把

$$I(p)=-\int p(x)\log p(x)dx \tag{33}$$

定义为分布的熵. 若零均值正态分布以 Γ_p 为协方差阵，则

$$p(x)=\frac{1}{(2\pi)^{\frac{p+1}{2}}(\det\Gamma_p)^{1/2}}\exp\left(-\frac{1}{2}x'\Gamma_p^{-1}x\right) \tag{34}$$

这时不难算出

$$I(p)=\log(2\pi e)^{\frac{p+1}{2}}(\det\Gamma_p)^{\frac{1}{2}}$$

这样，性质 1 表明 $f_1(\lambda)$ 所对应的协方差 $B_1(k)$ 是在满足(31)的条件下使(34)规定的正态分布密度具有最大熵，所以这个谱估计方法也称**最大熵谱估计法**.

性质 2 由(30)对应的 $f_1(\lambda)$ 是满足

$$\int_{-\pi}^{\pi}e^{in\lambda}f_1(\lambda)d\lambda=B_1(n)=\hat{B}(n),\quad |n|\leqslant p \tag{35}$$

的 $f_1(\lambda)$ 中使

$$I=\int_{-\pi}^{\pi}\log f_1(\lambda)d\lambda \tag{36}$$

为最大者.

[证明] f_1 满足(35)是定理 2 的结论. 若要满足(35)又使

(36)为最大，则由拉格朗日不定乘数法必有

$$\frac{\partial\left(\int_{-\pi}^{\pi}\log f_1 d\lambda-\sum_{n=-p}^{p}c_n\int_{-\pi}^{\pi}e^{in\lambda}f_1 d\lambda\right)}{\partial f_1}=0 \tag{37}$$

即

$$\frac{1}{f_1}=\sum_{n=-p}^{p}c_n e^{in\lambda}$$

$$f_1(\lambda)=\frac{1}{\sum_{n=-p}^{p}c_n e^{in\lambda}}$$

而(28)中的 $f_1(\lambda)$ 必可表为上述形式，所以它满足(37)，又由

$$\frac{\partial^2\left(\int_{-\pi}^{\pi}\log f_1 d\lambda-\sum_{n=-p}^{p}c_n\int_{-\pi}^{\pi}e^{in\lambda}f_1 d\lambda\right)}{\partial f_1^2}=-\frac{1}{f_1^2}<0$$

所以 $f_1(\lambda)$ 使(36)中的 f 极大.

在信息论中，也有把(36)规定的 I 定义作以 f_1 为谱密度的平稳过程的熵．所以这里介绍的用自回归模型拟合的谱密度估计法也就是满足(35)又使(36)规定的熵为最大的谱密度估计法.

上面我们介绍了用自回归模型拟合时间序列的问题．假若我们用有限参数模型来拟合实际资料，除了自回归模型外还可用滑动和模型或自回归滑动和混合模型，但它们的参数估计问题就要比自回归模型复杂得多，所以我们就不作介绍了，只举一个滑动和模型的例子来说明问题的复杂性.

若 $X_1(n)$ 满足滑动和模型

$$X_1(n)=\xi(n)+b\xi(n-1),\quad 0<b<1 \tag{38}$$

$\{\xi(n)\}$ 是方差为 σ^2 的白噪声，这时不难验证它的协方差函数是

$$B(n)=\begin{cases}\sigma^2(1+b^2), & n=0\\ b\sigma^2, & |n|=1\\ 0, & |n|\geq 2\end{cases} \tag{39}$$

由于
$$\frac{B(1)}{B(0)}=\frac{b}{1+b^2}$$

若限于 $|b|<1$, 则

$$b=\frac{B(0)-\sqrt{B^2(0)-4B^2(1)}}{2B(1)} \tag{40}$$

对于高阶的滑动和，由 $B(k)$ 求 b_j 的方程组就不是线性的了. 且一般不能求得象(38)一样的明显表示式而需要用迭代法来求解，即使求得了 b_j 的估计，进一步的理论分析表明，象(38)一样经 $\hat{B}(k)$ 求得的 $\hat{b}_j$，其方差比最大似然估计要大得多，但在此情况下的最大似然估计的求解是困难的. 譬如对(36)可作出比(38)规定的 $\hat{b}$ 更好的 b 的估计 $\tilde{b}$, 当 $b=\frac{1}{2}$ 时 $\tilde{b}$ 的方差只有 $\hat{b}$ 的 $\frac{1}{4}$ 左右, 所以对滑动和或自回归滑动和混合模型其参数估计问题就要复杂得多.

习　题

1. 若 $\{\xi(n)\}$ 为白噪声序列, $X(n)=\xi(n)+\xi(n-1)$. 试证 $H_0(X)=H_0(\xi)$ (提示: $\xi(0)=\lim\limits_{n\to\infty}\frac{1}{n}\sum\limits_{k=0}^{n}(-1)^k(n-k)X(k)$), 并由此求出用 $\{X(n), n\leqslant 0\}$ 线性预测 $X(1)$ 的均方误差.

2. 设 $\{X(n)\}$ 为平稳序列, 若用 $\{X(n), n\neq 0\}$ 对 $X(0)$ 内插, 令

$$\sigma^2=\inf_{c_j} E|X(0)-\sum_{n_j\neq 0}c_jX(n_j)|^2$$

试证 $\sigma^2>0$ 的充要条件是

$$\int_{-\pi}^{\pi}\frac{d\lambda}{F'(\lambda)}<\infty$$

其中 $F'(\lambda)$ 表 X 谱函数 $F(\lambda)$ 绝对连续部分的导数, 且

$$\sigma^2 = 2\pi\left(\int_{-\pi}^{\pi}\frac{d\lambda}{F'(\lambda)}\right)^{-1}$$

3. 若 $\{X(t)\}$ 为平稳过程，用 $aX(0)+bX(T)$ 来估计 $X(t)$ 时，试求出使均方误差为最小的 a, b 及相应的均方误差值.

4. 若 $\{X(t), -\infty<t<\infty\}$ 为平稳过程，若用 $aX(t)+bX'(t)$ 来估计 $X(t+\lambda)$ 时，试求出使均方误差为最小的 a, b 及相应的均方误差值. 并写出 $b=0$ 的充要条件.

5. 若 $X(t)=s(t)+n(t)$, $\{s(t)\}$, $\{n(t)\}$ 为相互独立平稳过程. 对固定的 λ, 用 $[aX(t)+bX(t-\lambda)]/\lambda$ 来估计 $s'(t)$, 试求出使均方误差最小的 a, b 及相应的均方误差值.

6. 若 $y=f(z, \bar{z})$, z 为复数. f 为二次连续可微函数，试证使 y 达到极小值的必要条件是

$$\frac{\partial f}{\partial z}=0, \qquad \frac{\partial f}{\partial \bar{z}}=0 \tag{*}$$

又若 $\overline{f(z_1, z_2)}=f(\bar{z}_2, \bar{z}_1)$, 则 (*) 及

$$\begin{pmatrix} \dfrac{\partial^2 f}{\partial z\partial \bar{z}} & \dfrac{\partial^2 f}{\partial z^2} \\ \dfrac{\partial^2 f}{\partial \bar{z}^2} & \dfrac{\partial^2 f}{\partial z\partial \bar{z}} \end{pmatrix}$$

为正定阵是使 y 在 z 达到极小的充分条件. 并对 $z_1, \bar{z}_1, \cdots, z_n, \bar{z}_n$ n 个复变数函数写出类似的达到极值的充分条件和必要条件.

7. 若

$$\mathscr{F}=\{f: f(\lambda)=2\pi\left|\sum_{j=0}^{p} c_j e^{-ij\lambda}\right|^{-2}, \int_{-\pi}^{\pi} f(\lambda)\,d\lambda=\hat{B}(0)\}$$

$$\mathscr{G}=\{g: \int_{-\pi}^{\pi} e^{ik\lambda} g(\lambda)\,d\lambda=\hat{B}(k), |k|\leqslant p\}$$

$$\rho(f, g)=-\int_{-\pi}^{\pi}[\log g/f-g/f+1]\,d\lambda$$

试证当 a_j, σ^2 按 (4.16)——(4.18) 决定, $\hat{f}(\lambda)$ 取为 (4.28) 时，满足对每个 $g\in\mathscr{G}$,

$$\rho(\hat{f}, g)=\min_{f\in\mathscr{F}}\rho(f, g)$$

8. 若 $\hat{B}(k)$ 由 (4.6) 规定, Γ 按 (4.8) 规定, 试证当 $\hat{B}(0)\neq 0$ 时，Γ 必是正定的.

第六章 随机模拟

本章将介绍随机模拟的基本思想与方法. 历史上第一个用这种方法的例子，就是有名的蒲丰投针问题(请参看第一册§1.4)，距今已有上百年的历史，但由于模拟试验工具的限制，一直没有得到发展. 直到本世纪四十年代出现了电子计算机后，这一方法才又受到重视. 随着要处理的问题日益复杂，传统的解析方法不能很好处理，以及计算机的迅猛发展. 这一方法也迅速得到发展，其结果不仅使概率论得到更广泛的应用，反过来，对概率论本身也产生日益增强的影响.

随机模拟，也有人称为蒙特卡洛(Monte-Carlo)或统计试验方法的. 近年来，国外也有人在此基础上，进一步提出计算概率论与统计的.

§1. 基本概念

一、积分计算的一种新方法

为了突出基本思想，我们来考察一个简单的积分计算问题(参看图1)：

$$I=\int_0^1 f(x)dx$$

也就是要计算由曲线 $y=f(x)$, x 轴、y 轴以及 $x=1$ 为边界的区域的面积. 显然，由于 x 轴和 y 轴的比例尺度可以任意改变，所以假定 $0\leqslant f(x)\leqslant 1$，即把区域限制在单位正方形内这一点并不影响问

题的一般性.

现在在单位正方形（$0\leqslant x\leqslant 1$, $0\leqslant y\leqslant 1$）中，随机地投掷一点，即它的两个坐标(x,y)都是$[0,1]$中均匀分布的随机变量，并且是相互独立的。容易看出，这一随机点落在所求区域中（即当$y\leqslant f(x)$）的概率p正好就是所求区域的面积，也就是所要计算的积分

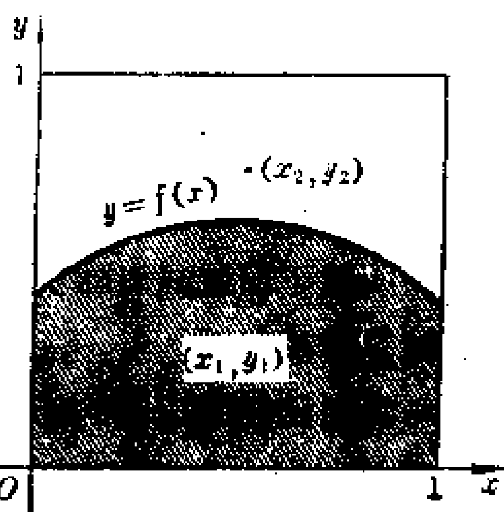


图 1

$$I=\int_0^1 f(x)dx=p$$

因此，如果重复上述试验n次，即随机地投掷n个点，假定有$m(\leqslant n)$个点落在所在区域之中，则根据大数定律，当n很大时，将有：

$$I=p\doteq\frac{m}{n}$$

这样，我们得到了经过多次统计试验而由频率估计概率以计算积分的新方法，这儿的“经过统计试验，由频率估计概率”正是随机模拟的基本思想所在。

二、随机模拟的基本思想

假如我们要求计算某个量x。如上所说，如果要用随机模拟的方法来计算，我们就要找到某一随机事件A，或者更一般地找到一个随机变量X，使

$$x=P\{A\}$$

或

$$x=E\{X\}$$

然后模拟事件A或随机变量X多次，并计算它出现的频率m/n或平均值

$$\overline{X}=\frac{X_1+X_2+\cdots+X_n}{n}$$

则当 n 充分大时, m/n 或 $\overline{X}$ 就可作为所求量 x 的近似值。

由于事件的概率可看作一个只取 0, 1 两个值的随机变量的均值，所以下面就只对随机变量及其均值这种形式来讨论这种方法的计算精度问题.

我们已经知道 X 的均值是 x, 现在假定它的方差存在，且记为 D, 于是根据中心极限定理可知, 当 n 充分大时,有:

$$P\left\{\mu<\frac{(X_1-x)+(X_2-x)+\cdots+(X_n-x)}{\sqrt{nD}}<\lambda\right\}$$

$$\doteq\frac{1}{\sqrt{2\pi}}\int_{\mu}^{\lambda}e^{-\frac{u^2}{2}}du$$

也就是

$$P\left\{\mu\cdot\sqrt{\frac{D}{n}}<\overline{X}-x<\lambda\cdot\sqrt{\frac{D}{n}}\right\}$$

$$\doteq\frac{1}{\sqrt{2\pi}}\int_{\mu}^{\lambda}e^{-\frac{u^2}{2}}du \tag{1}$$

特别取 $\mu=-3, \lambda=3$, 此时右边为 0.99, 所以可以说，以0.99 的概率有

$$|\overline{X}-x|<3\cdot\sqrt{\frac{D}{n}} \tag{2}$$

或者取 $\mu=-2, \lambda=2$, 此时(1)式右边为 0.95, 所以可以说, 以 0.95 的概率有

$$|\overline{X}-x|<2\cdot\sqrt{\frac{D}{n}} \tag{3}$$

等等.

在大多数实际问题中, 方差 D 一般是未知的, 这时可用它的估计来代替:

$$\hat{D}=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2$$

相应地，以 0.99 的概率有：

$$|\bar{X}-x|<3\cdot\sqrt{\frac{\hat{D}}{n}} \tag{2'}$$

和以 0.95 的概率有：

$$|\bar{X}-x|<2\cdot\sqrt{\frac{\hat{D}}{n}} \tag{3'}$$

等等。

由此可见，在随机模拟中，计算的精度是带有随机性质的，这是与一般计算方法中所说的精度有所不同的。

同时，由(2)或(3)都可以看出，计算的精度是与 $1/\sqrt{n}$ 成正比的，因而要提高精度，就必须增加模拟试验的次数。假如精度要提高 10 倍，试验次数就要提高 100 倍，这就限制了这种方法产生具有很高精度的解。

从(2)，(3)还可以看出，计算的精度还依赖于方差 D，它是与所选用的随机变量 X 有关的。因此，如何选取既满足 $E\{X\}=x$，又有较小方差 D 的 X，也是随机模拟中的一个重要问题。

文

三、进一步的例子

除了前面已经提到的蒲丰投针问题外，为了进一步掌握随机模拟的基本思想，下面再给出两个例子。

［例 1］ **高维积分** 现在来考虑高维积分的计算：

$$I=\int_0^1\cdots\int_0^1 f(x_1,x_2,\cdots,x_k)dx_1,\cdots,dx_k$$

且不妨假定 $0\leqslant f(x_1,\cdots,x_k)\leqslant 1$。容易看出，第一段中的方法可以毫无困难的搬过来，只是现在要在 $k+1$ 维空间中随机投点。确切的说，在 $k+1$ 维单位正方体 $(0\leqslant x_1\leqslant 1,\cdots,0\leqslant x_k\leqslant 1,0\leqslant y\leqslant 1)$ 中随机地投掷 n 个点，并求出满足 $y\leqslant f(x_1,x_2,\cdots,x_k)$ 的点的个数 m，

则 m/n 就是 I 的近似值.

这里我们可以指出随机模拟方法比一般数值积分方法的一些优点. 一个优点是, 对 $f(x_1, x_2, \cdots, x_k)$ 没有什么光滑性方面的限制, 计算的精度与 $f(x_1, \cdots, x_k)$的性质无关. 另一个更重要的优点是, 它的计算精度与维数 k 没有关系, 当然在具体实现时, 由于现在是在 $k+1$ 维空间中投点, 要产生 $k+1$ 个相互独立的[0, 1]中均匀分布的随机变量, 从而使计算量随着维数 k 而线性增长. 但在一般的数值积分方法中, 为了保证同样的精度, 其计算量一般要随着维数按幂次而增加.

[例 2] **随机面积的分布** 某实际问题可抽象为如下的问题: 试将一把筷子先垂直放置于桌上, 放手后, 筷子纷纷倒下, 要求这些筷子所张的面积的分布.

为了解决这一问题, 当然要将上面的叙述明确化. 我们假定开始时, 筷子都垂直放置在图 2 的各格子点上. 然后让各筷子随机倒下, 所谓"随机倒下", 是指筷子的底端不动, 而顶端落下后, 筷子与 x 轴的夹角 α 是$[0, 2\pi]$上的均匀分布(参看图 3). 再假定各筷子是相互独立地随机倒下, 而这些筷子所张的面积是指包含这些筷子端点的最小凸多边形的面积. 在这些假定下, 问题的意义

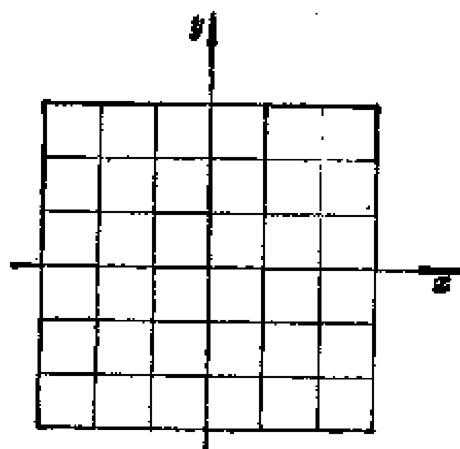


图 2

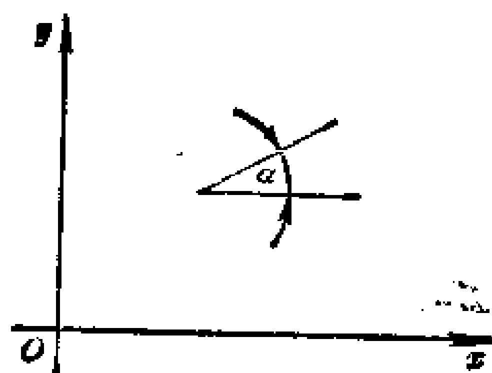


图 3

是唯一地确定了，但要想用传统的数学方法来解决它，求得某种解析的表达式，即使不是不可能的，也是极困难的.

我们在最后一节中，将要用随机模拟的方法来解决这个问题. 它的基本思想是在计算机中先模拟每根筷子的"随机倒下". 为此，在筷子的底端的位置(x_1, y_1)和筷子的长度l已知时，只要能产生一个$[0, 2\pi]$中均匀分布的随机变量的一个抽样值α，就可以求出筷子顶端在"随机倒下"后的位置(u_1, v_1):

$$\begin{aligned} u_1 &= x_1 + l\cos\alpha \\ v_1 &= y_1 + l\sin\alpha \end{aligned} \tag{4}$$

这就相当于模拟了一根筷子的"随机倒下". 类似地模拟完所有筷子的随机倒下后，即可求出这次试验的面积，重复上述的试验，可以求得一串面积的抽样值，然后运用已学过的数理统计方法，就可以对这种随机面积的分布进行种种的推断.

四、小结

1. 应用随机模拟方法解决问题的过程大体如下:

(i) 建立适当的概率模型(随机事件，随机变量等);

(ii) 进行模拟;

(iii) 对模拟结果进行统计处理(计算频率，平均值等)，以给出所求问题的解和解的精度的估计.

上述的(i)，随着问题的不同是不同的，即使同一问题也可以建立不同的概率模型，因此可以说是千变万化的. 目前也没有一般的方法，所以只能通过一些实例分析作些介绍，主要还有待读者自己去实践，这部分将在本章的最后一节中介绍.

上述的(ii)，是本章的重点. 我们将在下面的两节中分别介绍如何在计算机上产生各种分布的随机变量的抽样以及某些重要的随机过程的现实的方法.

上述的(iii)，在本节中已有所讨论，由于只是本书第二册有关

方法的应用，所以就不再讨论了.

2. 在应用随机模拟方法解决问题时，可能有两种状况，一种是原始的问题是非随机性的，如积分的计算等；另一种是原始的问题本身就是随机性的，如上段的例 2.

3. 随机模拟方法的优点是观念上比较简单，因而能用于求解复杂的问题；一般不需要什么连续性、可微性等限制，而且特别适用于高维的数学问题. 它的缺点是计算量大，计算的精度一般不很高，而且还是概率意义下的精度. 所以，随机模拟方法一般都用在无法用传统的物理试验或数学方法进行处理的复杂问题的求解中.

§2. 随机变量的模拟

本节将先介绍如何在计算机上产生[0, 1]上均匀分布的随机变量，有时也叫随机数的方法；然后介绍由[0, 1]上均匀分布的随机数构造任意分布的随机数的方法.

一、伪随机数 我们的任务是要在计算机上用一定的算法产生一串数$\{x_1, x_2, x_3, \cdots\}$，它们应当满足：

(i) 所得的数列在"统计性质"上，与从[0, 1]上的均匀分布抽样所得的子样相同，或至少相当地近似；

(ii) 产生一个随机数的运算量要比较少.

由于是用算法产生的，因而本质上是决定性的，再加上计算机字长有限，所以无论用什么算法产生的数列，在"统计特性"上都不可能完全与从均匀分布中抽样所得的子样完全相同，因而只能要求尽可能的近似. 这种在计算机用算法得到的"统计性质"上近似于[0, 1]上均匀分布的数，一般就称为**伪随机数**，以区别于真正从[0, 1]上均匀分布中抽样所取得的随机数. 我们的问题也就是要寻

找快速产生伪随机数的算法.

为了达到快速的要求，一般采用递推公式

$$x_{n+1}=f(x_n, x_{n-1}, \cdots, x_{n-k}) \tag{1}$$

在给定一组初值 $x_0, x_{-1}, x_{-2}, \cdots x_{-k}$ 后，即可逐步求出 $x_1, x_2, x_3, \cdots$.

目前最常用的方法是上述方法的一个特例:

$$y_{n+1}=a\cdot y_n+b \pmod M \quad 0\leqslant y_{n+1}<M$$

$$x_n=\frac{y_n}{M} \tag{2}$$

其中 a, b, M 以及初值 y_0 都是正整数. 容易看出 x_n 满足:

$$0\leqslant x_n<1 \tag{3}$$

这种方法称为**混合同余法**.

当 $b=0$ 时,

$$y_{n+1}=a\cdot y_n \pmod M$$

$$x_n=\frac{y_n}{M} \tag{4}$$

这时无需加法，它称为**乘同余法**.

当 $a=1, b\neq 0$ 时,

$$y_{n+1}=y_n+b \pmod M$$

$$x_n=\frac{y_n}{M} \tag{5}$$

这时无需乘法，它称为**加同余法**. 更一般的加同余法可以包含用前面几个 y 通过“加法”和“同余”来产生. 历史上比较有名的有下列的“非波那西”数列:

$$y_{n+1}=y_n+y_{n-1} \pmod M$$

$$x_n=\frac{y_n}{M} \tag{6}$$

虽然加同余法更快些，但所得数列的性质一般不如乘同余法和混合同余法，所以一般还是用乘同余法或混合同余法. 关于这

点，我们在下面还要提到.

在二进制的机器中，一般取 $M=2^W$，其中 W 是机器的字长，这时同余运算是十分简单的. 相应地，在十进制机器中，一般取 $M=10^W$. 为了进一步加快乘同余法的速度，可以取形如 2^l+1, 2^l+2^m+1, …这样的数作为 a，这时"乘 a"这样的运算法可用少量几个移位和加法来代替.

下面再对数列的性质作些讨论. 由于我们要求所得的数列 $\{x_1, x_2, \cdots, x_n\}$ 尽可能具有从[0, 1]上均匀分布随机抽样所得子样的各种统计特性. 为此，最简单的方法是运用第二册中各种假设检验的方法对下列原假设进行检验:

H_0: 母体分布是[0, 1]上的均匀分布

下面我们列出两个最常用的检验法.

1. **频率检验** 用于检验经验频率与理论频率(即概率) 是否有显著性差异.

将[0, 1)区间分为 k 个等长度的区间:

$$\left[0, \frac{1}{k}\right), \left[\frac{1}{k}, \frac{2}{k}\right), \cdots, \left[\frac{k-1}{k}, 1\right)$$

用 f_i 表示 $\{x_1, x_2, \cdots, x_n\}$ 中落在区间 $\left[\frac{i-1}{k}, \frac{i}{k}\right)$ 中的个数，则在 H_0 假设下 (即在 $\{x_1, x_2, \cdots, x_n\}$ 是[0, 1) 上均匀分布的随机子样的假设下)，由分布的 χ^2 检验法(第二册定理 3.4.1)可知，下列统计量

$$\chi_1^2=\sum_{i=1}^{k}\frac{\left(f_i-n\cdot\frac{1}{k}\right)^2}{n\cdot\frac{1}{k}}$$

$$=\frac{k}{n}\sum_{i=1}^{k}\left(f_i-\frac{n}{k}\right)^2$$

是渐近 $\chi^2(k-1)$ 分布的. 由此即可进行显著性检验.

进一步，如重复上述做法，也就是产生 m 组长度为 n 的随机

数：$\{x_1^{(j)}, x_2^{(j)}, \cdots, x_n^{(j)}\}, 1\leqslant j\leqslant m$，对每一组，可以计算出一个相应的 χ_1^2，共 m 个：

$$\chi_1^2(1), \chi_1^2(2), \cdots, \chi_1^2(m)$$

然后可以用 χ^2 检验法对上述数据进行这组数据是否服从 $\chi^2(k-1)$ 分布随机抽得的子样的假设进行检验.

2. **独立性检验** 仿前，将 $[0, 1)$ 分成 k 个等长度的区间，用 f_{ij} 表示 $\{x_1, x_2, x_3, \cdots, x_n\}$ 中，本身落在 $\left[\frac{i-1}{k}, \frac{i}{k}\right)$ 中，且其后面一数恰好落在 $\left[\frac{j-1}{k}, \frac{j}{k}\right)$ 中的数的个数. 计算

$$\chi_2^2=\frac{k^2}{n}\sum_{i,j=1}^{k}\left(f_{ij}-\frac{n}{k^2}\right)^2$$

则在 H_0 假设下，可以证明（参看下引文献及其后所引文献①）：$\chi_2^2-\chi_1^2$ 是渐近 $\chi^2(k^2-k)$ 分布的，由此即可进行显著性检验.

类似地，如产生一组 $\{x_1^{(l)}, x_2^{(l)}, \cdots, x_n^{(l)}\}, 1\leqslant l\leqslant m$，则可求得一组

$$\chi_2^2(1)-\chi_1^2(1), \chi_2^2(2)-\chi_1^2(2), \cdots, \chi_2^2(m)-\chi_1^2(m)$$

对它还可以进行是否自从 $\chi^2(k^2-k)$ 分布中随机抽样的假设进行检验.

一般，如果所得数列能通过上述的或其它的统计检验，我们就认为它的随机性比较的好. 但是从观念上讲，用算法产生的数列，显然是决定性的，而不是随机性的. 所以从六十年代开始，有些人企图从别的角度对由算法产生的数列的“伪随机性”进行更严格的讨论，也取得了一定的成功. 下面对这方面作些简要的介绍：

设 $X=\{x_1, x_2, x_3, \cdots\}$ 是一个数列，且所有的 x_n 都在 0, 1 之间：

$$0\leqslant x_n\leqslant 1$$

---

① 参看 T. E. Hull 等，“Mixed congruential random number generators”，J. ACM Vol. 11(1964)31——40 及其所引文献.

用 $\nu_N(a)$表示$\{x_1, x_2, \cdots, x_N\}$中满足:

$$0 \leqslant x_n < a \leqslant 1$$

的 x 的个数,如果

$$\lim_{N\to\infty}\frac{\nu_N(a)}{N}=a$$

对一切 $0\leqslant a\leqslant 1$ 都成立,我们就说数列X是**等分布的**。

进一步,对任意的正整数 k 和数列 X,我们来考虑由相继 k 个数组成的数组构成的序列:

$$(x_{n+1}, x_{n+2}, \cdots, x_{n+k}), \quad n=1, 2, 3, \cdots$$

如果用 $\nu_N(a_1, a_2, \cdots, a_k)$表示

$$(x_{n+1}, x_{n+2}, \cdots, x_{n+k}), \quad n=1, 2, 3, \cdots, N$$

中满足:

$$0\leqslant x_{n+i}<a_i\leqslant 1, \quad i=1, 2, \cdots, k$$

的数组个数,当对一切[0, 1]中的 $a_1, a_2, \cdots, a_k$ 都有:

$$\lim_{N\to\infty}\frac{\nu_N(a_1, a_2, \cdots, a_k)}{N}=a_1, a_2, \cdots, a_k$$

时,则称数列X是**k 维等分布的**. 容易看出, 上面提到的等分布,相当于一维等分布.

如果数列 X,对任意 k, 都是 k 维等分布的, 则它就称为**完全等分布的**.

由于上述定义中,并不要求数列X是随机的,但又较好地体现了[0, 1]上均匀分布, 以及独立抽样的要求, 因而就比较成功地对非随机数列的"伪随机性"给出了确切的描述.

作为例子,我们来看一个有名的外耳序列:

$$x_n=\{n\alpha\}, \quad n=1, 2, 3, \cdots$$

其中{ }表示小数部分,例如{2.305}=0.305. 我们来证明当 α 是无理数时,它是等分布的.

考虑下列三角和

$$T_k=\frac{1}{N}\sum_{n=1}^{N}e^{2\pi ik\{n\alpha\}}=\frac{1}{N}\sum_{n=1}^{N}e^{2\pi ikn\alpha}$$

其中 $i=\sqrt{-1}$. 容易看出 $T_0=1$, 而对 $k\neq0$, 将上述几何级数求和可得

$$T_k=\frac{1}{N}\frac{1-e^{2\pi ik(N+1)\alpha}}{1-e^{2\pi ik\alpha}}=\frac{e^{-\pi ikN\alpha}}{N}\cdot\frac{\sin\pi k(N+1)\alpha}{\sin\pi k\alpha}$$

于是有

$$|T_k|\leqslant\frac{1}{N}\frac{1}{\sin\pi k\alpha},\quad k\neq0$$

由于 α 是无理数, 所以 $k\alpha$ 不可能是整数, 因而 $\sin\pi k\alpha\neq0$. 于是由上式可得

$$\lim_{N\to\infty}T_k=0,\quad k\neq0$$

现在设 $P(x)=\sum_k a_k e^{2\pi ikx}$ 为任意有限阶三角多项式, 则有:

$$\begin{aligned}\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}P(\{n\alpha\})&=\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}\sum_k a_k e^{2\pi ik\{n\alpha\}}\\&=\lim_{N\to\infty}\sum_k a_kT_k=a_0T_0=a_0\end{aligned}$$

另一方面, 显然有:

$$\int_0^1 P(x)dx=a_0$$

所以对任意有限阶三角多项式 $P(x)$, 有

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}P(\{n\alpha\})=\int_0^1 P(x)dx$$

由于[0, 1] 上的任意连续函数都可以用三角多项式来一致逼近, 因此上式对一般的连续函数也是成立的.

对任意给定的[0, 1]中的一个数 a, 我们引入如下两个连续函

数 $f^+(x)$和 $f^-(x)$:

$$f^+(x)=\begin{cases}1, & 0\leqslant x\leqslant a\\ \text{在 } 0, 1 \text{ 之间}, & a\leqslant x\leqslant a+e\\ 0, & a+e\leqslant x\leqslant 1\end{cases}$$

$$f^-(x)=\begin{cases}1, & 0\leqslant x\leqslant a-e\\ \text{在 } 0, 1 \text{ 之间}, & a-e\leqslant x\leqslant a\\ 0, & a\leqslant x\leqslant 1\end{cases}$$

其中 e 为一小的正数. 不难看出

$$a\leqslant\int_0^1 f^+(x)dx\leqslant a+e$$

$$a-e\leqslant\int_0^1 f^-(x)dx\leqslant a$$

以及

$$\sum_{n=1}^{N} f^-(\{n\alpha\})\leqslant\nu_N(a)\leqslant\sum_{n=1}^{N} f^+(\{n\alpha\})$$

其中 $\nu_N(a)$表示 $\{n\alpha\}, n=1, 2, \cdots, N$ 中满足: $0\leqslant\{n\alpha\}<a$ 的$\{n\alpha\}$的个数. 综合上述各式, 即可得

$$\lim_{N\to\infty}\frac{\nu_N(a)}{N}=a$$

这就证明了数列 $x_n=\{n\alpha\}$是等分布的. 还可以证明①, 它一般不是 $k(k>1)$维等分布的, 当然更不是完全等分布的.

但是, 可以证明②, 下列序列

$$x_n=\{\theta^n\}$$

对几乎所有的 $\theta>1$ 是完全等分布的, 其中 $\{\cdot\}$ 仍然表示小数部分. 事实上, 要上述的 x_n 是 k 维等分布的, 就必须使 θ 不是低于

① 参看 W. Freiberger 等, "A course in computational probability and statistics", Springer-Verlag, 1971, 第 13 页.

② 参看 D. E. Knuth, "Construction of a random sequence", BIT 5(1965) 246——250.

k 阶的代数数，即不能是某一低于 k 阶的代数多项式的根；于是，要上述的 x_n 是完全等分布的，θ 必须不是代数数，即必须是超越数①.

显然，序列$\{n\alpha\}$相应于加同余法，而序列$\{\theta^n\}$相应于乘同余法，由此也可看出，乘同余法所得的序列，其性质较为良好.

在计算机上产生伪随机数时，还要碰到计算机字长有限带来的问题. 这方面的讨论是十分困难的，下面我们只限于讨论由此产生的一个问题——周期性问题.

容易看出，用混合同余法(2)产生数列时，只要一旦发生：

$$y_1 = y_{1+p}$$

那么就一直有

$$y_{1+k} = y_{1+p+k}, \quad k=1, 2, \cdots$$

也就是说，从 $p+1$ 项开始，由混合同余法产生的数列就开始重复，也就是出现了周期现象. 但由于由混合同余法产生的 y_n 总满足 $0 \leqslant y_n < M$，因而是有限的，所以总存在某个正整数 p，使 $y_1 = y_{1+p}$ 成立，从而出现周期现象. 使上式成立的最小正整数 p，一般称为它的**周期**. 显然周期越长，不同的 y_n 就越多，从而 x_n 也越多，也就越可能符合均匀分布或等分布的要求，因而也就越好. 对给定的 M（在二进制机器中，它总是取为 2^W，在十进制机器中，总是取为 10^W，其中 W 由机器的字长所决定）. 显然有 $p \leqslant M$，但并不是总能找到适当的系数 a, b, x_0，使 $p=M$ 的.

下面我们不加证明地给出有关周期的一些结果②：

（i）如 $M=2^W$，$W \geqslant 3$；$b=0$（即乘同余法），则当且仅当 $a = \pm 3 \pmod 8$时，产生的序列有最大的周期，它是 $2^{W-2} = \frac{M}{4}$，即没有

① 见上页注②

② 请参看 J. H. Halton, "A Restrospective and Prospective Survey of the Monte-Carlo Method", SIAM Review 12(1970)1——63.

达到M.

（ii） 如 $M=10^W$, $W\geqslant 4$; $b=0$, 此时能达到的最大周期为 $5\times 10^{W-2}=\frac{M}{20}$, 同样没有达到$M$. 能达到这种周期的$a$至少有如下一些:

$$a=\pm 3, \pm 13, \pm 27, \pm 37, \pm 53, \pm 67, \pm 77, \pm 83 \pmod{200}$$

在所有上述两种情况，都要求初值 y_0 与M互质，一般就取 $y_0=1$.

(iii) 如果不要求 $b=0$, 即如果用混合同余法，则最大周期可达M, 但要求：(a) b 与M互质；(b)对M的每一质因子 p，有 $b=1 \pmod p$; (c)如M能被 4 整除，则还要求 $b=1 \pmod 4$. 这时初值可以随意取，一般就取为 0.

现在再回到本段一开始的问题上去：我们的目的是要寻找一种既快速，又能产生出性质好的伪随机数的算法. 为了快速，目前最常用的就是混合同余法和乘同余法，而为了使其中的乘法更快，往往要求其中的参数 a 有某种特殊的形式，如 $a=2^t+1$ 等. 而为了要有长的周期，能通过各种统计检验等，又对 a, b 等参数的选择加上了另外的一些限制. 上面提到的关于周期方面的一些结果等，对于选择参数都是十分有用的. 但可惜至今尚无寻求最佳参数a, b, x_0 的一般的方法. 寻求好的参数仍然需要在理论的指导下，通过多次试验才能解决.

下面是两组较好的参数的例子，更多的材料，读者可参看前引的文献:

（i） $W=35$, $M=2^{35}$

$$a=2^s+1, \quad 5\leqslant s\leqslant 11$$
$$b=1$$
$$y_0=0$$

（ii） $W=35$, $M=2^{35}$

$a=2^s+3,\quad 6\leqslant s\leqslant 30$

$b=0$, 即乘同余法

$y_0=1$

最后，我们再简略地讨论一下由于用伪随机数代替随机数而带来的影响问题. 因为我们用的虽然是伪随机数，但由于我们总要求它们能通过多种统计检验，因此一般说来，问题是不大的. 因而用伪随机数所得的结果也是可信的. 有趣的是，在有些特定的场合，用伪随机数还会得到比用随机数更好的结果. 下面是这样的一个例子.

考虑[0,1] 上的有三阶连续导数的函数 $f(x)$，我们要计算下列积分:

$$I=\int_0^1 f(x)dx$$

我们已经知道，如用蒙特卡洛方法来求，当用真的随机数时，其误差是与$\frac{1}{\sqrt{N}}$成正比的，其中N是试验的次数. 但现在如果用前面提到过的伪随机数

$$x_n=\{n\alpha\},\quad n=1,2,\cdots$$

并用(参看本章后面 § 4 第一段)

$$\hat{I}_N=\frac{1}{N}\sum_{n=1}^{N}f(x_n)=\frac{1}{N}\sum_{n=1}^{N}f(\{n\alpha\})$$

来估计，则可以证明①

$$|\hat{I}_N-I|\leqslant C\cdot\frac{1}{N}$$

其中C是某一常数. 因此，其精度反而更高.

由此可以引出一连串有待解决的问题，如

① 见前引 W. Freiberger 等的书，第 24 页到第 44 页.

(i) 多维积分时，有否类似的现象？

(ii) 用别的伪随机数，又会怎样？对任意给定的$f(x)$，有否最佳的伪随机数列？

(iii) 如进一步考虑计算机的有限字长效应，又会怎样？

二、离散型随机变量的模拟 设Z是离散型随机变量：

$$P\{Z=a_i\}=p_i$$

其中$p_i\geqslant 0$, $\sum_i p_i=1$. 现在的问题是要在计算机上产生Z的随机子样.

设$x_1, x_2, x_3, \cdots$是$[0, 1]$上均匀分布的随机子样(产生它们的近似方法已在上段中讨论过了)，令

$$z_k=\begin{cases} a_1, \text{当 } 0\leqslant x_k<p_1 \\ a_2, \text{当 } p_1\leqslant x_k<p_1+p_2 \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ a_i, \text{当 } p_1+\cdots+p_{i-1}\leqslant x_k<p_1+\cdots+p_i \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \end{cases} \qquad k=1, 2, \cdots$$

则容易看出，由此造出的$z_1, z_2, z_3, \cdots$正是Z的随机子样. 事实上，我们有：

$$P\{Z_k=a_i\}=P\{p_1+\cdots p_{i-1}\leqslant X_k<p_1+\cdots+p_i\}$$

由于假定X是$[0, 1]$上的均匀分布，所以上式右边为

$$(p_1+\cdots+p_i)-(p_1+\cdots+p_{i-1})=p_i$$

因此

$$P\{Z_k=a_i\}=p_i$$

这正是我们所需要的.

三、一般随机变量的模拟 设Z是一般的随机变量，其分布函数为$F(u)$，即

$$P\{Z<u\}=F(u)$$

于是如果 $F(u)$ 连续且严格单调上升，则易知

$$X = F(Z) \tag{7}$$

是[0, 1]中的均匀分布，因此，如 X 是[0, 1]上均匀分布的随机变量，则

$$Z = F^{-1}(X) \tag{8}$$

便是以 $F(u)$ 为分布函数的随机变量，其中 $F^{-1}(v)$ 是 $F(u)$ 的反函数.

于是，如已得一组[0, 1]上均匀分布的子样 $\{x_1, x_2, x_3, \cdots\}$，即可由

$$z_k = F^{-1}(x_k), \quad k = 1, 2, \cdots$$

而求得一组以 $F(u)$ 为分布函数的随机变量的子样. 所以可以说[0, 1]上均匀分布随机数是产生其它任意分布随机数的基石，有效地产生[0, 1]上均匀分布随机数是提高随机模拟方法质量的基础.

[例] **指数分布** 它的分布密度是

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geqslant 0 \\ 0, & x < 0 \end{cases}$$

相应的分布函数是

$$F(x) = \begin{cases} 1 - e^{-\lambda x}, & x \geqslant 0 \\ 0, & x < 0 \end{cases}$$

由(7)知，如 Z 为指数分布的随机变量，则

$$X = 1 - e^{-\lambda Z}$$

为[0, 1)上均匀分布的随机变量，而

$$Z = -\frac{1}{\lambda}\ln(1-X)$$

又因当 X 是[0, 1]上均匀分布随机变量时，$1-X$ 也是[0, 1]上均匀分布，所以

$$\tilde{Z} = -\frac{1}{\lambda}\ln X \tag{9}$$

即为指数分布的随机变量.

顺便指出，由于k个独立的指数分布之和的分布是指数为k的爱尔兰分布E_k，所以如果$X_1,\cdots,X_k$是k个相互独立的[0,1]上的均匀分布的随机变量，则

$$Z=-\frac{1}{\lambda}\ln X_1-\frac{1}{\lambda}\ln X_2-\cdots-\frac{1}{\lambda}\ln X_k$$
$$=-\frac{1}{\lambda}\ln(X_1\cdot X_2\cdots X_k)$$

就是E_k分布的随机变量.

四、舍选法

这个方法的实质是从许多均匀分布的随机数中选出一部分，使选出的那部分有指定的分布，它可用于产生任意有界的随机变量Z.

设Z为在(a,b)中取值的随机变量，其密度为$f(z)$，并选常数λ，使$\lambda f(z)\leqslant 1$，$z\in(a,b)$. 又设X_1,X_2为相互独立的[0,1]上均匀分布的随机变量，于是

$$Y_1=a+(b-a)X_1$$

是(a,b)中的均匀分布的随机变量. 现在如有$X_2\leqslant\lambda f(Y_1)$，则令

$$Z=Y_1$$

否则舍去这一组X_1,X_2(或者形式上定义$Z=\infty$)，并重新产生一对[0,1]上的均匀分布的采样值，直到$X_2\leqslant\lambda f(Y_1)$满足而产生一个真正的$Z$为止. 此时，有

$$\begin{aligned}&P\{Z<z\mid X_2\leqslant\lambda f(Y_1)\}\\&=P\{Z<z,X_2\leqslant\lambda f(Y_1)\}/P\{X_2\leqslant\lambda f(Y_1)\}\\&=P\{Y_1<z,X_2\leqslant\lambda f(Y_1)\}/P\{X_2\leqslant\lambda f(Y_1)\}\\&=\frac{\frac{1}{b-a}\int_a^z\lambda f(y_1)dy_1}{\frac{1}{b-a}\int_a^b\lambda f(y_1)dy_1}=\int_a^z f(y_1)dy_1\end{aligned}$$

其中利用了$\int_a^b f(y_1)dy_1=1$. 于是可知，按上述方法产生的 Z 确实有分布密度 $f(z)$.

一个应用舍选法的基本思想但有所变化的例子如下：

[例] $[0,2\pi]$**上均匀分布的正弦和余弦** 它当然可以这样产生：先产生二个$[0,1]$上的均匀分布随机变量 X_1, X_2，然后令

$$Z_1=\sin(2\pi X_1)$$
$$Z_2=\cos(2\pi X_2)$$

即可. 但由于要用正弦和余弦函数，运算量一般比较大，利用舍选法可以进行如下：产生一对$[0,1]$上的均匀分布的随机变量 X_1 和 X_2. 如 $X_1^2+X_2^2>1$，则舍去，重新产生一对；如 $X_1^2+X_2^2\leqslant 1$，令

$$Z_1=\pm 2X_1X_2/(X_1^2+X_2^2)$$
$$Z_2=(X_1^2-X_2^2)/(X_1^2+X_2^2)$$

则可以证明 Z_1 和 Z_2 就是$[0,2\pi]$上均匀分布的正弦和余弦，其中 Z_1 的符号是随机选取的.

五、近似方法

近似方法是很多的，这里只介绍用逐段逼近分布的近似方法，因为它比较方便而且普遍适用.

假设要求得到的分布函数为 $F(z)$. 首先将 $(-\infty,\infty)$分为 n 个区间：

$$(-\infty,a_1],\ (a_1,a_2],\ \cdots,\ (a_{n-1},\infty)$$

然后在各区间中，分别用较简单的函数 $F_k(z)$来"近似"$F(z)$：

$$F_k(z)\doteq\frac{F(z)-F(a_{k-1})}{p_k},\quad 当\ z\in(a_{k-1},a_k]$$

其中

$$p_k=F(a_k)-F(a_{k-1}),\quad k=1,2,\cdots,n$$

且假定 $a_0=-\infty, a_n=+\infty$. 于是可用下列方法产生以 $F(z)$ 为近似分布函数的随机变量 $\tilde{Z}$.

1. 产生[0, 1]上的均匀分布的随机变量 X_1，根据它的采样值决定下一步在哪个小区间上继续进行，即如

$$p_1+\cdots+p_{k-1}\leqslant X_1<p_1+\cdots+p_k$$

则下一步在第 k 个小区间 $(a_{k-1}, a_k]$ 上进行。

2. 产生[0, 1]上的均匀分布 X_2，并令

$$\tilde{Z}=F_k^{-1}(X_2)$$

最简单的是"线性近似"，也就是取

$$F_k(z)=\frac{z-a_{k-1}}{a_k-a_{k-1}}, \quad a_{k-1}\leqslant z\leqslant a_k$$

相应的

$$\tilde{Z}=a_{k-1}+(a_k-a_{k-1})X_2$$

六、正态分布 由于正态分布经常出现，这里给出了两个产生正态分布的方法.

首先，由于任意均值为 μ，方差为 σ^2 的正态分布 $N(\mu, \sigma^2)$ 可由 $N(0, 1)$ 的随机变量 Z，经下列线性变换而得到

$$X=\mu+\sigma\cdot Z$$

所以下面只讨论产生 $N(0, 1)$ 分布的随机变量的方法.

变换方法 设 X_1, X_2 是[0, 1] 上均匀分布的相互独立的随机变量，则可以证明：

$$\begin{aligned} Z_1&=\sqrt{-2\ln X_1}\cos 2\pi X_2 \\ Z_2&=\sqrt{-2\ln X_1}\sin 2\pi X_2 \end{aligned} \tag{10}$$

Z_1, Z_2 是相互独立的 $N(0, 1)$ 分布的随机变量. 事实上，由(10)可解出 X_1, X_2：

$$\begin{aligned} X_1&=\exp\left\{-\frac{1}{2}(Z_1^2+Z_2^2)\right\} \\ X_2&=\frac{1}{2\pi}\operatorname{tg}^{-1}\left(\frac{Z_2}{Z_1}\right) \end{aligned} \tag{11}$$

于是有

$$\frac{\partial X_1}{\partial Z_1}=e^{-\frac{1}{2}(Z_1^2+Z_2^2)}\cdot(-Z_1)$$

$$\frac{\partial X_1}{\partial Z_2}=e^{-\frac{1}{2}(Z_1^2+Z_2^2)}\cdot(-Z_2)$$

$$\frac{\partial X_2}{\partial Z_1}=\frac{1}{2\pi}\frac{1}{1+\left(\frac{Z_2}{Z_1}\right)^2}\left(-\frac{Z_2}{Z_1^2}\right)$$

$$\frac{\partial X_2}{\partial Z_2}=\frac{1}{2\pi}\frac{1}{1+\left(\frac{Z_2}{Z_1}\right)^2}\frac{1}{Z_1}$$

和

$$\begin{vmatrix}\frac{\partial X_1}{\partial Z_1} & \frac{\partial X_1}{\partial Z_2}\\ \frac{\partial X_2}{\partial Z_1} & \frac{\partial X_2}{\partial Z_2}\end{vmatrix}=\begin{vmatrix}e^{-\frac{1}{2}(Z_1^2+Z_2^2)}(-Z_1) & e^{-\frac{1}{2}(Z_1^2+Z_2^2)}(-Z_2)\\ \frac{1}{2\pi}\frac{1}{1+\left(\frac{Z_2}{Z_1}\right)^2}\left(-\frac{Z_2}{Z_1^2}\right) & \frac{1}{2\pi}\frac{1}{1+\left(\frac{Z_2}{Z_1}\right)^2}\cdot\frac{1}{Z_1}\end{vmatrix}$$

$$=-\frac{1}{2\pi}e^{-\frac{1}{2}(Z_1^2+Z_2^2)}$$

由于 X_1, X_2 是相互独立的[0, 1] 上的均匀分布，所以 Z_1, Z_2 的联合分布密度为:

$$\frac{1}{2\pi}e^{-\frac{1}{2}(Z_1^2+Z_2^2)}=\frac{1}{\sqrt{2\pi}}e^{-\frac{Z_1^2}{2}}\cdot\frac{1}{\sqrt{2\pi}}e^{-\frac{Z_2^2}{2}}$$

即 Z_1, Z_2 是相互独立的 $N(0, 1)$分布的随机变量.

近似方法 上述方法由于要用到开方，取对数，以及正弦，余弦等，其计算量一般是比较大的. 在精度要求不高时，可以采用下面介绍的近似方法.

设 $X_1, X_2, \cdots, X_n$ 是[0, 1] 上均匀分布的相互独立的随机变量，易知其均值和方差分别为$\frac{1}{2}$和$\frac{1}{12}$，则根据中心极限定理，

$$Z=\frac{\sum_{i=1}^{n}\left(X_i-\frac{1}{2}\right)}{n\cdot\frac{1}{12}}$$

在n较大时，其分布近似于$N(0,1)$分布。

一般取$n=12$，于是有

$$Z=\sum_{i=1}^{12}\left(X_i-\frac{1}{2}\right)=X_1+\cdots+X_{12}-6$$

$$=X_1+\cdots+X_6-(1-X_7)-\cdots-(1-X_{12})$$

注意到X_i和$1-X_i$都是$[0,1]$上的均匀分布，所以可取

$$Z=X_1+\cdots+X_6-X_7-\cdots-X_{12}$$

它也是近似于$N(0,1)$分布的．这一做法，在一般情况下是够用的，但由于$|Z|\leqslant 6$，所以在必须考虑Z取特大或特小数值这一情况时，还必须采用其它更为精确的方法。

七、多维分布

以二维分布为例进行讨论．对一般的二维分布(Z_1,Z_2)，如其分布密度为$f(z_1,z_2)$，则可将它写为

$$f(z_1,z_2)=f(z_1)f(z_2|z_1)$$

其中$f(z_1)$为Z_1的分布密度，$f(z_2|z_1)$为Z_2关于Z_1的条件分布密度．于是可用前面介绍的方法，先根据$f(z_1)$产生Z_1，然后根据$f(z_2|z_1)$产生Z_2，所得的(z_1,z_2)即为以$f(z_1,z_2)$为联合分布密度的二维随机变量的抽样。

对多维正态随机变量，有更简便的方法．设$(Z_1,\cdots,Z_n)$是均值为$(\mu_1,\cdots,\mu_n)$，协方差矩阵为：

$$\Sigma=\begin{bmatrix}\sigma_{11} & \sigma_{12} & \cdots & \sigma_{1n}\\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2n}\\ \cdots & \cdots & \cdots & \cdots\\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_{nn}\end{bmatrix}$$

的n维正态分布的多维随机变量，其中

$$\sigma_{ij}=E\{(Z_i-\mu_i)(Z_j-\mu_j)\}$$

由于Σ是正定对称的，所以由矩阵论可知，存在下三角矩阵

C，使

$$\Sigma = C \cdot C' \tag{12}$$

其中 C' 表示C的转置. 如果已经应用前面介绍的方法求得 n 个相互独立的$N(0,1)$的随机变量 $Y_1, Y_2, \cdots, Y_n$，则可以验证

$$Z = \mu + CY \tag{13}$$

其中

$$Z = \begin{bmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_n \end{bmatrix}, \quad \mu = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{bmatrix}, \quad Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}$$

是以 μ 为均值，Σ 为协方差阵的 n 维正态随机变量，事实上，我们有

$$E\{Z\} = \mu + C \cdot E\{Y\} = \mu + C \cdot 0 = \mu$$

和

$$\begin{aligned} & E\{(Z_i - \mu_i)(Z_j - \mu_j)\} = E\{(Z - \mu)(Z - \mu)'\} \\ & = E\{CY(CY)'\} = E\{CYY'C'\} \\ & = CE\{YY'\}C' = C(E\{Y_iY_j\})C' \\ & = C \begin{pmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{pmatrix} C' = CC' = \Sigma \end{aligned}$$

顺便指出，由(12) 式从 Σ 求C的计算是很方便的. 下面对 $n=3$ 的情形写出有关算式，一般场合是完全类似的. 我们有

$$\begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{bmatrix} = \begin{bmatrix} c_{11} & 0 & 0 \\ c_{21} & c_{22} & 0 \\ c_{31} & c_{32} & c_{33} \end{bmatrix} \begin{bmatrix} c_{11} & c_{12} & c_{13} \\ 0 & c_{22} & c_{23} \\ 0 & 0 & c_{33} \end{bmatrix}$$

而且其中有 $c_{ij}=c_{ji}$. 根据矩阵乘法的定义，对第一行元素有：

$$\sigma_{11}=c_{11}^2$$

$$\sigma_{12}=c_{11}c_{12}$$

$$\sigma_{13}=c_{11}c_{13}$$

由此解得

$$c_{11}=\sqrt{\sigma_{11}}$$

$$c_{12}=c_{21}=\sigma_{12}/\sqrt{\sigma_{11}}$$

$$c_{13}=c_{31}=\sigma_{13}/\sqrt{\sigma_{11}}$$

对第二行上半部的元素有：

$$\begin{cases}\sigma_{22}=c_{21}c_{12}+c_{22}c_{22}=c_{12}^2+c_{22}^2=\dfrac{\sigma_{12}^2}{\sigma_{11}}+c_{22}^2\\[2ex] \sigma_{23}=c_{21}c_{13}+c_{22}c_{23}=c_{12}c_{13}+c_{22}c_{23}=\dfrac{\sigma_{12}\sigma_{13}}{\sigma_{11}}+c_{22}c_{33}\end{cases}$$

由此解得：

$$\begin{cases}c_{22}=\sqrt{\sigma_{22}-\dfrac{\sigma_{12}^2}{\sigma_{11}}}\\[2ex] c_{23}=\left(\sigma_{23}-\dfrac{\sigma_{12}\sigma_{13}}{\sigma_{11}}\right)\Big/\sqrt{\sigma_{22}-\dfrac{\sigma_{12}^2}{\sigma_{11}}}\end{cases}$$

对第三行上半部元素，有：

$$\sigma_{33}=c_{31}c_{13}+c_{32}c_{23}+c_{33}c_{33}$$
$$=c_{13}^2+c_{23}^2+c_{33}^2$$

由此解得

$$c_{33}=\sqrt{\sigma_{33}-\frac{\sigma_{13}^2}{\sigma_{11}}-\frac{\left(\sigma_{23}-\dfrac{\sigma_{12}\sigma_{13}}{\sigma_{11}}\right)^2}{\left(\sigma_{22}-\dfrac{\sigma_{12}^2}{\sigma_{11}}\right)}}$$

特别当

$$\sigma_{ij}=\sigma^2 e^{-\alpha|i-j|}$$

其中 $\sigma>0, \alpha>0$ 为常数时，可以验证：

$$c_{i1}=\sigma e^{-(i-1)\alpha}$$

$$c_{ij}=\sigma e^{-(i-j)\alpha}\sqrt{1-e^{-2\alpha}},\qquad 2\leqslant j\leqslant i$$

此时 $Z'=CY$ 化为:

$$z_1'=c_{11}y_1=\sigma y_1$$

$$\begin{aligned}z_2'&=c_{21}y_1+c_{22}y_2\\&=\sigma e^{-\alpha}y_1+\sigma\sqrt{1-e^{-2\alpha}}\,y_2\\&=e^{-\alpha}z_1'+\sigma\sqrt{1-e^{-2\alpha}}\,y_2\end{aligned}$$

$$\begin{aligned}z_3'&=c_{31}y_1+c_{32}y_2+c_{33}y_3\\&=\sigma e^{-2\alpha}y_1+\sigma e^{-\alpha}\sqrt{1-e^{-2\alpha}}\,y_2+\sigma\sqrt{1-e^{-2\alpha}}\,y_3\\&=e^{-\alpha}(\sigma e^{-\alpha}y_1+\sigma\sqrt{1-e^{-2\alpha}}\,y_2)+\sigma\sqrt{1-e^{-2\alpha}}\,y_3\\&=e^{-\alpha}z_2'+\sigma\sqrt{1-e^{-2\alpha}}\,y_3,\end{aligned}$$

一般有

$$z_k'=e^{-\alpha}z_{k-1}'+\sigma\sqrt{1-e^{-2\alpha}}\,y_k,\quad 2\leqslant k\leqslant n$$

由此可得产生均值为$(\mu_1,\cdots,\mu_n)$，协方差阵为$(\sigma^2\cdot e^{-\alpha|i-j|})$的 n 维正态随机变量的递推算法如下:

1. 产生 n 个独立的 $N(0,1)$分布随机变量 $y_1,y_2,\cdots,y_n$;
2. 按下列递推关系计算 z_k':

$$z_1'=\sigma\cdot y_1$$

$$z_k'=e^{-\alpha}z_{k-1}'+\sigma\sqrt{1-e^{-2\alpha}}\,y_k,\ 2\leqslant k\leqslant n$$

3. 计算 z_k:

$$z_k=\mu_k+z_k'$$

这一结果，我们在下一节中要用到。

§3. 随机过程的模拟

一、平稳正态过程

设 $Z(t)$是均值为 0，协方差函数为 $B(\tau)$的平稳正态过程，我

们要产生它的现实．在计算机上，当然只能模拟有限个值：

$$Z(k\Delta t),\quad 1\leqslant k\leqslant n$$

其中 Δt 为一正的常数．

实际中最常见的两种 $B(\tau)$ 是

$$B_1(\tau)=\sigma^2e^{-\alpha|\tau|},\qquad \sigma>0,\alpha>0\ \text{为常数}\qquad (1)$$

和

$$B_2(\tau)=\sigma^2e^{-\alpha|\tau|}\cos\beta\tau,\ \sigma>0,\quad \alpha>0,\beta>0\ \text{为常数}\qquad (2)$$

下面先介绍实现上述二种协方差函数的方法．

当协方差函数为 $B_1(\tau)$ 时，易知 $\{Z(k\Delta t),\ 1\leqslant k\leqslant n\}$ 的协方差阵为：

$$(\sigma^2e^{-\alpha|i-j|\Delta t})$$

于是应用上节第七段的结果，可得如下的方法：

1．产生 n 个相互独立的 $N(0,1)$ 分布的 $y_1,\cdots,y_n$；

2．计算：

$$z_1=\sigma y_1$$

$$z_k=e^{-\alpha\Delta t}z_{k-1}+\sigma\sqrt{1-e^{-2\alpha\Delta t}}\,y_k,\qquad 2\leqslant k\leqslant n$$

有了产生 $B_1(\tau)$ 的方法后，即可产生 $B_2(\tau)$，其方法如下：

1．产生两个独立的平稳正态过程 $Z_1(t)$，$Z_2(t)$，使其均值为0，协方差函数都是 $\sigma^2\cdot e^{-\alpha|\tau|}$．

2．计算：

$$Z(t)=Z_1(t)\cos\beta t+Z_2(t)\sin\beta t$$

可以证明，它的协方差函数就是 $B_2(\tau)$．事实上，考虑到 $Z_1(t)$ 和 $Z_2(t)$ 独立和它们的均值为零，我们有

$$\begin{aligned}&E\{Z(t)Z(t+\tau)\}\\&=E\{(Z_1(t)\cos\beta t+Z_2(t)\sin\beta t)\\&\quad\cdot(Z_1(t+\tau)\cos\beta(t+\tau)+Z_2(t+\tau)\sin\beta(t+\tau))\}\\&=E\{Z_1(t)Z_1(t+\tau)\cos\beta t\cos\beta(t+\tau)\}\end{aligned}$$

$$+E\{Z_2(t)Z_2(t+\tau)\sin\beta t\sin\beta(t+\tau)\}$$

$$=\sigma^2\cdot e^{-\alpha|\tau|}\cos\beta t\cos\beta(t+\tau)$$

$$+\sigma^2\cdot e^{-\alpha|\tau|}\sin\beta t\sin\beta(t+\tau)$$

$$=\sigma^2\cdot e^{-\alpha|\tau|}[\cos\beta t\cos\beta(t+\tau)$$

$$+\sin\beta t\sin\beta(t+\tau)]$$

$$=\sigma^2\cdot e^{-\alpha|\tau|}\cos\beta\tau$$

对于一般的平稳正态过程，

$$Z(k\Delta t),\qquad 1\leqslant k\leqslant n$$

是一个 n 维的正态随机变量，它当然可以应用上一节第七段中的方法产生，但 n 如果很大时，计算量一般是很大的，所以就要采用近似的方法.

方法的基本思想是，对给定的 $B(\tau)$ 和 Δt，先用 § 5.4 中的方法拟合一个 p 阶的自回归模型：

$$Z(k\Delta t)\doteq Z_k$$

$$=-a_0Z_{k-1}-\cdots-a_{p-1}Z_{k-p}+\sigma^2Y_k,\qquad k\geqslant 1 \tag{3}$$

其中 $\{Y_k,\ k\geqslant 1\}$ 是相互独立的 $N(0,1)$ 的随机变量. 在选定 p 后，根据 § 5.4 可知，$\{a_j\}$ 和 σ^2 由下式决定：

$$\begin{bmatrix} B(0) & B(\Delta t) & \cdots & B((p-1)\Delta t) \\ B(\Delta t) & B(0) & \cdots & B((p-2)\Delta t) \\ \cdots\cdots & \cdots\cdots & \cdots & \cdots\cdots \\ B((p-1)\Delta t) & B((p-2)\Delta t) & \cdots & B(0) \end{bmatrix}\begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_{p-1} \end{bmatrix}$$

$$=\begin{bmatrix} B(\Delta t) \\ B(2\Delta t) \\ \vdots \\ B(p\Delta t) \end{bmatrix} \tag{4}$$

$$\sigma^2=\sqrt{B(0)+a_0B(\Delta t)+\cdots+a_{p-1}B(p\Delta t)} \tag{5}$$

于是对于给定的 $B(\tau)$ 和 Δt，在选定一个 p（也可用 § 5. 4 中

的方法)之后，即可先用(3)解出 $a_0, a_1, \cdots, a_{p-1}$，然后用(4)求得 σ^2. 如果再有了 p 个初值 $Z_0, Z_{-1}, \cdots, Z_{1-p}$，即可由一组相互独立的 $N(0, 1)$ 分布的 Y_k，由(3)式产生出一组以 $B(\tau)$ 为近似协方差函数的平稳正态过程的现实 $z_1 \doteq Z(\Delta t), z_2 \doteq Z(2\Delta t), \cdots$.

初值的选取有两种方法:

一种是任取 p 个值作为初值，但将生成的序列 z_k 的前面某一段舍去不用，以此来减弱初值的影响.

另一种方法是用上一节第七段中的方法，生成一个 p 维的正态分布，使其协方差阵为:

$$\begin{bmatrix} B(0) & B(\Delta t) & \cdots & B((p-1)\Delta t) \\ B(\Delta t) & B(0) & \cdots & B((p-2)\Delta t) \\ \cdots & \cdots & \cdots & \cdots \\ B((p-1)\Delta t) & B((p-2)\Delta t) & \cdots & B(0) \end{bmatrix}$$

并用这 p 个值作为初始，这种方法比上一种方法要好，但更化费时间.

至于 p 的选取，还没有一般的方法，主要通过试验决定.

二、马尔可夫链

设所要模拟的马尔可夫链 $\{Z_n\}$ 是齐次的，且有转移概率

$$P = (p_{ij}), \qquad i, j = 0, 1, 2, \cdots$$

和初始分布 $\{p_i\}, i = 0, 1, 2, \cdots$.

用上节第二段的方法，即可产生一个以 $\{p_i\}$ 为分布的随机变量 Z_0，设其采样值为 z_0，则

$$\{p_{z_0, j}\}, \qquad j = 0, 1, 2, \cdots$$

关于 j 为一个概率分布，于是仍用上节第二段的方法，即可产生一个以它为分布的随机变量 Z_1. 再设其采样值为 z_1，则重复上述步骤，即可产生 $Z_2, Z_3, \cdots$.

用这种方法产生的 $z_0, z_1, z_2, \cdots$ 就是以 $\{p_i\}$ 为初始分布，以

$P=(p_{ij})$为一步转移概率的齐次马尔可夫链的现实.

显然用这种方法也可以产生非齐次的马尔可夫链.

现在来考虑模拟时间连续状态离散且有限的齐次马尔可夫链 $Z(t)$. 假设其初始分布为 $p_i=P\{Z(0)=i\}$, 其转移概率为

$$P(t)=(p_{ij}(t)),\qquad i,j=0,1,2,\cdots,n$$

相应的Q矩阵为:

$$Q=(q_{ij})$$

引入

$$q_i=-q_{ii}$$

$$p_{ij}=\begin{cases}0, & i=j\\ q_{ij}/q_i, & i\neq j\end{cases}$$

则在有限状态时，总有

$$\sum_{j=0}^{n}p_{ij}=1$$

即(p_{ij})为一随机矩阵. 此时过程可如下模拟.

1. 根据分布$\{p_i\}$产生 y_0; 然后产生一个参数为 q_{y_0} 的指数分布的随机数 τ_1.

2. 根据$\{p_{y_0,j}\}$，产生 y_1; 然后产生一个参数为 q_{y_1} 的指数分布的随机数 τ_2; 根据$\{p_{y_1,j}\}$, 产生 y_2; 然后产生一个参数为 q_{y_2} 的指数分布的随机数 τ_3; 如此重复，即可产生一组 y_3, τ_3; y_4, τ_4; $\cdots$.

3. 对任意非负的 t, 如 $t<\tau_1$, 则令

$$Z(t)=y_0$$

否则如

$$\tau_1+\tau_2+\cdots+\tau_k\leqslant t<\tau_1+\tau_2+\cdots+\tau_{k+1},\quad k\geqslant1$$

则令

$$Z(t)=y_k,\quad k\geqslant1$$

关于这样造出的$Z(t)$就是所需要的过程①.

一般的马尔可夫过程经适当的离散近似后，均能化为上述两种情况而进行模拟. 同时，由于计算机的离散特性，实际上总是要做这种或那种离散近似的.

§4. 一 些 应 用

为了突出处理问题的基本思想，以及篇幅的限制，这里举出的例子是经过简化了的. 但掌握了这些之后，处理更复杂的问题时，也不会碰到原则上的困难.

一、计算积分

用随机模拟方法来计算积分，主要应用于高维的场合，这是由随机模拟方法的特点所决定的. 但由于基本思想在一维时和高维时是一样的，所以我们还是主要就一维的情形来进行讨论.

设要计算积分：

$$I=\int_0^1 f(x)dx$$

我们在§1中已经给出一种把I当作某一随机事件概率的随机模拟算法. 现在我们再给出另一种计算I的随机模拟算法.

设X是$[0,1]$上的均匀分布的随机变量，我们来考察随机变量$f(X)$. 容易看出，它的数学期望是

$$E\{f(X)\}=\int_0^1 f(x)dx=I$$

也就是说，我们要计算的I，恰好是随机变量$f(X)$的数学期望，由此利用大数法则即可给出计算I的另一种随机模拟算法：

1. 产生n个相互独立的$[0,1]$上的均匀分布的随机数X_1,

① 证明可参看：K. L. Chung, "Markov Chains with stationary transition probabilities"一书的第II部分.

$X_2, \cdots, X_n$;

2. 计算:

$$\frac{1}{n}\sum_{i=1}^{n} f(X_i)$$

并把它作为 I 的近似值.

这一算法推广到多维是直接的. 事实上, 假定要计算

$$I_2=\int_0^1\int_0^1 f(x, y)dxdy$$

则只要产生 $2n$ 个相互独立的 [0, 1]上均匀分布的随机变量 X_1, Y_1; X_2, Y_2; …; X_n, Y_n. 然后计算

$$\frac{1}{n}\sum_{i=1}^{n} f(X_i, Y_i)$$

并用它作为 I_2 的近似值.

这种计算积分的算法一般称为**期望值算法或平均值算法**, 而 § 1 中的算法则一般称为**掷点算法**. 下面我们来比较一下这两种算法.

在期望值算法中, 我们利用随机变量 $Z^{(1)}=f(X)$ 和

$$I=E\{f(X)\}$$

由 § 1 的(1), (2), (3)等式可知, 用

$$\hat{I}_1=\frac{1}{n}\sum_{i=1}^{n} f(X_i)$$

来估计 I 的精度与 $f(X)$ 的方差有关, 方差越大, 精度越差. 容易求得 $f(X)$ 的方差为:

$$D_1=\int_0^1 (f(x)-I)^2 dx$$

在掷点算法中, 当 $0\leqslant f(x)\leqslant 1$ 时, 我们实质上利用了随机变量

$$Z^{(2)}=\begin{cases}1, & \text{若 } Y\leqslant f(X)\\ 0, & \text{若 } Y>f(X)\end{cases}$$

和

$$I=E\{Z^{(2)}\}$$

同样，由 § 1 可知，用

$$\hat{I}_2=\frac{1}{n}\sum_{i=1}^{n}Z_i^{(2)}=\frac{m}{n}$$

(其中 m 为 n 次试验中，发生 $Y\leqslant f(X)$ 的次数)来估计 I 的精度与 $Z^{(2)}$ 的方差有关，而 $Z^{(2)}$ 是二点分布，易知其方差为

$$D_2=I\cdot(1-I)$$

这样我们有

$$\begin{aligned}D_2-D_1&=I\cdot(1-I)-\int_0^1(f(x)-I)^2dx\\&=I-I^2-\int_0^1 f^2(x)dx+I^2\\&=\int_0^1 f(x)dx-\int_0^1 f^2(x)dx\\&=\int_0^1 f(x)(1-f(x))dx\geqslant 0\end{aligned}$$

最后一个不等式是因为我们假定了 $0\leqslant f(x)\leqslant 1$. 由此可见 $D_2\geqslant D_1$，也就是说，掷点算法的精度不如期望值算法的好. 此外，在掷点算法中，每产生一个 $Z^{(2)}$，需要产生两个[0, 1]上均匀分布的随机数，计算一个 $f(X)$，再作一次比较：是否 $Y\leqslant f(X)$；而在期望值算法中，每产生一个 $Z^{(1)}$，只要产生一个[0, 1]上均匀分布的随机数，计算一个 $f(x)$，所以就一次模拟计算而言，也是期望值运算的运算量更少. 因此，计算积分时总是用期望值算法的.

如果两种算法，一种在精度上较好，另一种在运算量上较好，这时为了比较这两种算法，就要引进方法效率比的概念.

设 T_1, T_2 和 D_1, D_2 分别是两种不同算法在一次模拟试验中的平均运算量(或计算时间)和方差，则由 § 1 的(1)，(2)，(3)等式可

知，为了达到给定的精度，所需的总的计算时间是与$T \cdot D$成正比的，因此定义

$$C=\frac{T_2D_2}{T_1D_1}$$

为算法 I 和算法 II 的效率比. $C<1$，则算法 II 的效率高于算法 I.

关于如何寻求效率高的算法的问题，一般与具体的问题都有密切的关系，这儿就不再介绍了.

二、随机面积的分布

我们来继续讨论 § 1 第三段中的例 2. 在没有高速数字计算机之前，要想研究与比较复杂的概率模型有关的分布问题，几乎是不可能的. 于是往往只能人为地进行许多简化. 这样在为复杂的实际问题建立概率模型时，就给人们带来了很多的限制. 数字计算机的出现和发展，为人们的研究复杂的概率模型提供了工具，从而使人们在构造模型时，获得更大的自由来运用更符合实际的复杂模型. 这也是计算机对概率统计研究的重要影响之一. 这里讨论的随机面积分布和下面一个例子中关于排队模型的讨论，都是来说明这一点的.

如 § 1 中所说，我们解决这一复杂的随机面积分布问题的整个步骤如下：

1. 根据给定的条件，决定m根“筷子”的底端的位置，在我们这儿恰好分布在正方形的格子点上，只要给定格子的边长h后，确定这些点的位置当然是很容易做到的.

2. 对底端确定的“筷子”决定其随机落下后，顶端的位置. 这一点可以根据那儿的(4)式由一个$(0,2\pi)$中的均匀分布来决定，或者用 § 2 第四段舍选法的例子中介绍的方法直接产生$\cos\alpha$和$\sin\alpha$来决定都可以，并将这种工作对m根“筷子”都进行一遍. 这

样除了原来的m个底端外，又产生了m个顶端，共计$2m$个点.

3. 要决定上面得的$2m$个点所张的凸多边形的面积. 为确定起见，记这$2m$个点为(s_i, t_i)，$1\leqslant i\leqslant 2m$. 计算面积的一种较好的算法如下:

(i) 决定最左边的点: P_0. 为此，只要找出$\{s_i\}$，$1\leqslant i\leqslant 2m$中对应于最小值的i，这当然是容易让计算机做到的.

(ii) 求P_1，使$\overrightarrow{P_0P_1}$与Y轴的夹角最小. 夹角最小现在等价于夹角的余弦最大，而P_0与任一点的联线与Y轴的夹角的余弦都是可以用(s_i, t_i)，$1\leqslant i\leqslant 2m$表示出来的，因此寻找P_1的事也是容易让计算机做到的.

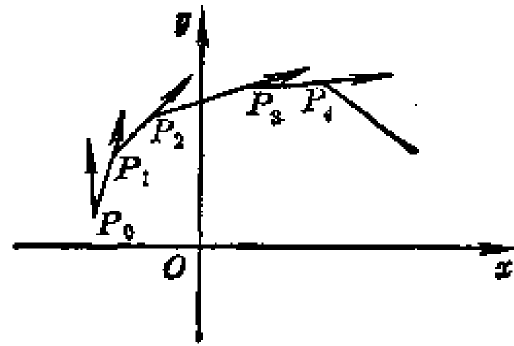


图 1

文

(iii) 求P_2使$\overrightarrow{P_1P_2}$与$\overrightarrow{P_0P_1}$的夹角最小；求P_3使$\overrightarrow{P_2P_3}$与$\overrightarrow{P_1P_2}$的夹角最小；……直到P_k与P_0重合为止. 在此过程中，逐步求出$\triangle P_0P_1P_2$, $\triangle P_0P_2P_3$, $\triangle P_0P_3P_4$, …的面积，并把它们相加，最后的和数就是这$2m$个点所张凸多边形的面积.

我们可以粗略地估计一下上述算法的运算量. 由于每决定一个边界点，都基本上要考察一遍所有的$2m$个点，因此其运算量是与$2m$，即点数成正比的，而边界点的个数，则可以认为与$\sqrt{2m}$，即点数的平方根成正比，于是总的运算量应与点数$2m$的$3/2$次方$(2m)^{3/2}$成正比.

4. 重复上述1, 2, 3, n次，即可求得随机面积的n个抽样值$A_1, A_2, \cdots, A_n$. 总的运算量是$n\cdot m^{3/2}$成正比的，有了这些数据，我们就可以对随机面积进行种种统计推断了.

我们曾就正方形格子点的布局，对各种特定的h——格子边长，$m=m_1^2$，其中m_1是每边上的筷子数，l——筷子的长度，n——

试验重复次数，进行计算，结果倒是相当简单的，关于随机面积的分布为正态分布的原假设能通过 χ^2 检验.

显然，上述算法稍作变化，就可以对不同的布局，以及"筷子"随机倒下时，整个筷子还伴有某种随机跳动等情况进行分析，而不会引起什么新的困难，这些都是普通的解析方法所不可能做到的.

可以这么说，在计算机广泛应用之前，一般总要得到解答的明显形式才算解决了问题. 而如果只是象我们上面那样，仅仅给出一个能在有限步中获得近似解答的算法，往往并不认为已经解决了问题. 但在计算机应用日益广泛以及要处理的问题越来越复杂时，可以认为算法给出的解答也是解决了问题. 只要算法是好的，它甚至是比有些解析解更为有用. 事实上，有些方程的解，要用很复杂的积分式才能表示解的富利埃变换或拉普拉斯变换等，其用处是有一定局限性的.

三、系统的模拟

我们来介绍一个单服务员排队系统的模拟，所述方法推广到更复杂的系统并没有什么原则上的困难，这点是与解析方法十分不同的.

考虑一个单服务员的排队系统. 相继两顾客来到的时间间隔是相互独立的随机变量 τ，其分布为:

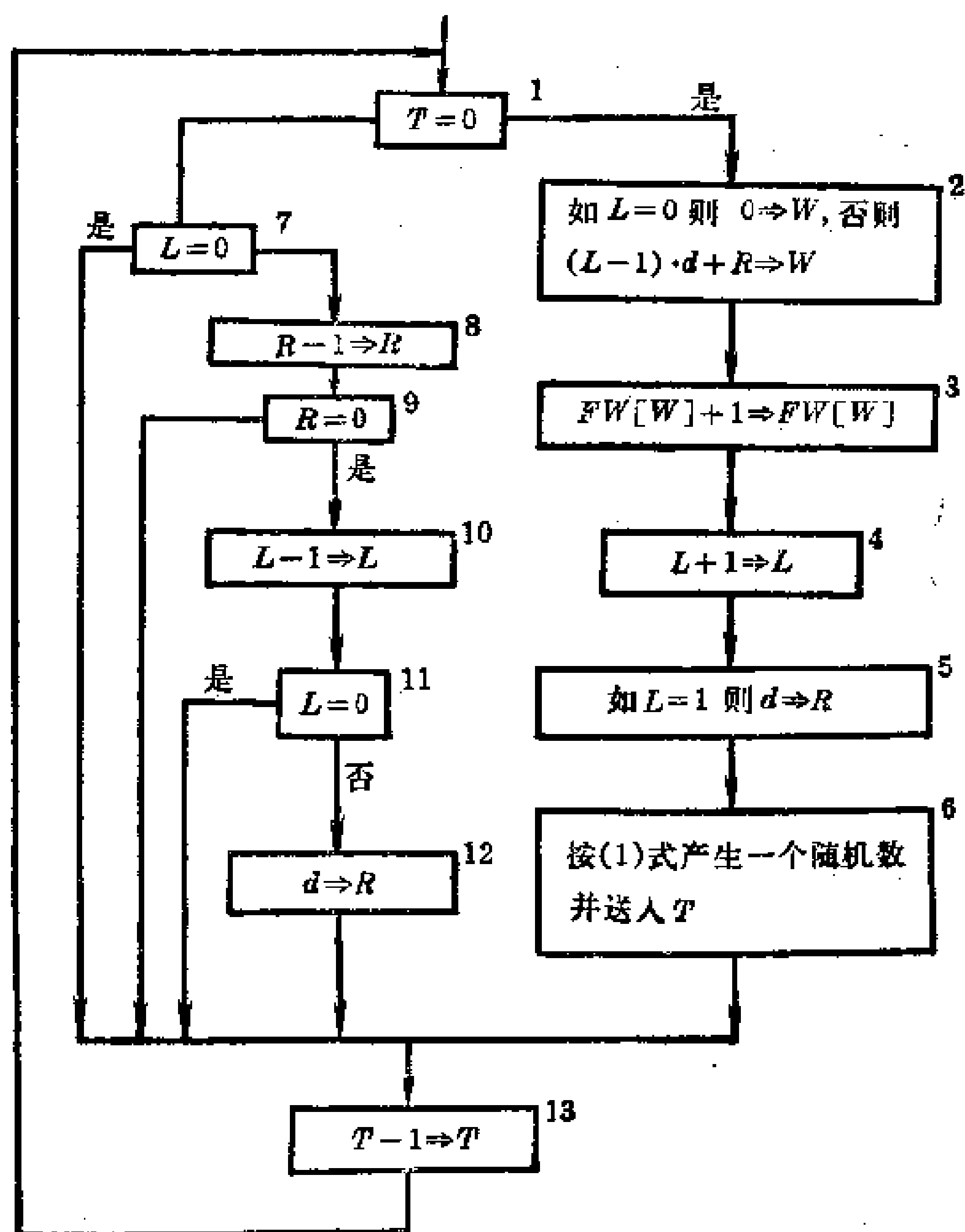
$$P\{\tau=i\}=p_i, \qquad 1\leqslant i\leqslant k \tag{1}$$

每个顾客的服务时间假定是整的常数 d. 排队规则是先来先服务. 我们要研究这一排队系统的等待时间分布.

我们现在每隔一单位时间，模拟一次系统的变化，并记录逐个顾客等待时间取不同值的频数，其总框图如下:

先对框图中所用的记号作些说明:

T: 表示当前时刻与下一次顾客来到时刻之间的时间间隔; 特别 $T=0$，即表示当前时刻将有一顾客来到.



L：表示当前时刻的队长，这里的队长理解为包括正在服务的顾客和正在排队的顾客的总数.

W：表示当前来到的顾客的等待时间. 如这顾客来到时，无人在服务，即 $L=0$，则等待时间 $W=0$，否则其等待时间应等于正在服务的顾客的剩余的服务时间 R 加上正在排队的 $L-1$ 个人的服务时间：$(L-1)d$.

R：表示当前时刻正在接受服务的顾客的剩余的服务时间；特别$R=0$表示一个顾客完成服务离去，从而可使队长减1.

$FW[0:M]$：是一个数组，其元素$FW[j](1\leqslant j\leqslant M)$表示等待时间恰好为$j$的顾客的个数，其中$M$为一充分大的数，视具体情况而定.

下面再对有些框作些简要的说明：

第1框：判断当前时刻是否有顾客来到，如有转到第2框，否则转第7框.

第2框：计算当前来到顾客的等待时间.

第3框：对等待时间进行计数.

第4框：由于当前时刻来到顾客，而使队长加1.

第5框：表示当前时刻来到的顾客立即得到服务，因此相应的刚开始接受服务的顾客的剩余服务时间为d.

第10框：由$R=0$来，也就是说，一个顾客恰好完成服务，从而使队长减1.

第12框：由$L\neq 0$来，表示队中再有顾客时，刚开始接受服务的顾客的剩余服务时间为d.

其余的框是自明的.

整个框图表示每隔一个单位时间后，刻划排队系统的一些量的变化．为了简化，框图中，没有画出置初态的框．显然在开始时，应令$T=L=W=R=0$，以及$FW[j]=0, 1\leqslant j\leqslant M$. 此外在框图中也没有说明整个框图要执行几次.

整个模拟过程结束后，在数组FW中就是我们所要的结果，它们分别是等待时间恰好为0, 1, 2, …的频数，除以总的模拟次数即为频率，也就是近似的概率.